

With CD-ROM
and 3-D viewer

Image Interpretation in Geology

STEVE DRURY

THIRD EDITION

b

Blackwell
Science

Image Interpretation in Geology

Since the first edition was published in 1987 *Image Interpretation in Geology* has established itself as essential reading for earth science, environmental science and physical geography students studying the geological applications of remote sensing and image interpretation.

The book describes the fundamentals of remote data capture and image processing, their practical limitations, and new techniques such as digital radar imaging and hyperspectral data analysis. Geological applications such as mapping, mineral exploration, and geohazards are illustrated by numerous black-and-white photographs and a colour plate section. A list of URLs encourages exploration of some of the excellent remote image resources on the internet.

New to the 3rd Edition is a CD-ROM (Mac and PC format) which contains an image gallery (with accompanying spectacles for viewing in 3-D), exercises, and TNTLite. This software, a fully-functional version of Microlmages Inc's TNTMips, encourages students to experience at first-hand the immense power of modern image processing and interpretation software, and will allow lecturers to devise a wide range of practical exercises to support their courses.



Copublished in the United States and
Canada with Blackwell Science, Boston

ISBN 0-632-05408-5
(USA and Canada only)

www.blackwell-science.com

T
nelson
thornes

www.nelsonthornes.com

ISBN 0-7487-6499-2



9 780748 764990

Image Interpretation in Geology

Steve Drury

*The Open University
Keynes*

Third Edition



b

Blackwell
Science

Copyright © Stephen A. Drury 2001

The right of Stephen A. Drury to be identified as author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the publisher or under licence from the Copyright Licensing Agency Limited, of 90 Tottenham Court Road, London W1T 4LP.

Any person who commits any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

First published in 1987 by Allen and Unwin
Second edition published in 1993 by Chapman and Hall
Third edition published in 2001:

Published in the UK by:
Nelson Thornes Ltd
Delta Place
27 Bath Road
CHELTENHAM
GL53 7TH
United Kingdom

Published in the USA and Canada by:
Blackwell Science Inc.
Commerce Place
350 Main Street
Malden, MA 01248-5018
USA

01 02 03 04 05 / 10 9 8 7 6 5 4 3 2 1

A catalogue record for this book is available from the British Library

ISBN 0 7487 6499 2

Library of Congress Cataloging-in-publication Data

Drury, S. A. (Stephen A.), 1946 –
Image Interpretation in Geology / Steve Drury.-3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0 632 05408 5
1. Geology-Remote Sensing. I. Title
QE33.2.R4 D78 2001
550'.22'2-dc21

For further information on Nelson Thornes visit our website:
www.nelsonthornes.com

For further information on Blackwell Science Inc. visit our website:
www.blackwell-science.com

The inclusion of TNTlite in the accompanying CD-ROM to *Image Interpretation in Geology* (Third edition) does not imply any endorsement of the software by either the publishers or the author.

Set by Graphicraft Limited, Hong Kong

Printed and bound in China by L. Rex

Contents

Preface, vii

Acknowledgements, x

Electromagnetic Radiation and Materials, 1

1.1 The nature of electromagnetic radiation, 1

1.2 The generation of electromagnetic radiation, 2

1.3 Matter and electromagnetic radiation, 4

Associated resources on the CD-ROM, 14

Further reading, 14

Human Vision, 16

2.1 The eye and visual cortex, 16

2.2 Spatial resolving power, 19

2.3 Seeing brightness, 24

2.4 Producing, seeing and representing colour, 26

2.5 Perception of depth, 29

2.6 Dangerous illusions, 30

Further reading, 31

3 How Data Are Collected, 32

3.1 Photography, 32

3.2 Vidicon cameras, 37

3.3 Line-scanning systems, 38

3.4 Pushbroom systems, 42

3.5 Microwave imaging, 43

3.6 Imaging spectrometers, 49

3.7 Gamma-ray spectrometers, 49

3.8 A short history of remote sensing, 50

3.9 Airborne data, 51

3.10 Basic characteristics of orbiting satellites, 53

3.11 Data from staffed spacecraft, 55

3.12 Data from unstaffed spacecraft, 57

3.13 Future prospects, 65

Associated resources on the CD-ROM, 67

Further reading, 67

4 Photogeology, 68

4.1 Destructional landforms, 69

4.2 The recognition of rock types, 76

4.3 Stratigraphic relationships, 90

4.4 Structural relationships, 95

4.5 Superficial deposits and constructional landforms, 107

Associated resources on the CD-ROM, 120

Further reading, 120

5 Digital Image Processing, 122

5.1 The image histogram, 123

5.2 Contrast stretching, 125

5.3 Spatial-frequency filtering, 133

5.4 Data reduction, 138

5.5 Band ratioing, 143

5.6 Pattern recognition, 145

Associated resources on the CD-ROM, 157

Further reading, 157

6 Thermal Images, 160

6.1 What a thermal image shows, 160

6.2 Qualitative interpretation of thermal images, 164

6.3 Semiquantitative analysis, 169

6.4 Multispectral thermal data, 172

Associated resources on the CD-ROM, 174

Further reading, 174

7 Radar Remote Sensing, 176

7.1 Interactions between radar and surface materials, 176

7.2 Interpretation of radar images, 180

7.3 Radar interferometry, 203

Associated resources on the CD-ROM, 204

Further reading, 204

8 Non-Image Data and Geographical Information Systems, 206

8.1 Forms of non-image data, 207

8.2 Non-image data in raster format, 208

8.3 Data analysis in geographical information systems, 221

8.4 Concluding note, 224

Associated resources on the CD-ROM, 225

Further reading, 225

9 Geological Applications of Image Data, 227

9.1 Geomorphology, 228

9.2 Geological mapping, 231

9.3 Exploration, 233

9.4 Engineering applications, 241

9.5 Geochemical hazards, 243

Further reading, 244

Appendix A Stereometry, 249

Further reading, 252

Appendix B Image Correction, 253

B.1 Geometric rectification, 253

B.2 Replacing dropped lines, 255

B.3 De-striping, 255

B.4 Removal of random noise, 257

Associated resources on the CD-ROM, 257

Further reading, 258

Appendix C The CD-ROM Resources, 259

Contents of the CD-ROM, 259

Installing resources, 260

Working with the IIG resources, 262

Getting help, 262

Appendix D Sources of Image Data, 264

Image atlases, 264

Websites, 264

Glossary, 266

Index, 279

Colour plates fall between pp. 148 and 149.

Preface

The first two editions of *Image Interpretation in Geology* won a wide readership among undergraduate and professional geologists, since 1987, but advances in the technology of remote sensing and its application demand some updating. In addition, there are always better ways of expressing concepts, so I have revised the style in most of the chapters. I have replaced several of the images with more instructive ones, and the further reading extends to 1999. Most important, I have added a CD-ROM that I hope will supplement and extend the usefulness of the book. In its new form, *Image Interpretation in Geology* transcends the original textbook to become potentially a complete, distance-learning course on remote sensing, image processing and digital mapping for geologists.

Thanks to MicroImages of Lincoln, Nebraska, the CD includes the students' version of their professional mapping and image processing system, TNTmips. This package, called TNTlite, is a unique example of highly sophisticated freeware, in that it is equally as functional as the professional version, limited only in terms of the maximum usable image size and with export from TNTmips format to other formats disabled. The CD includes both Windows and MacOS versions, so that almost everyone with access to a modern desktop or portable computer will be able to learn essential skills to a high level of sophistication. The package contains the full and comprehensive TNTmips reference manual and a series of Adobe Acrobat format, *Getting Started* tutorials, which cover all aspects of the system's functionality. Moreover, I have added 11 exercises that focus on all the geological applications of remote sensing covered by the text, which are based on a variety of image and nonimage data, mainly covering an area of stunning geology in the semiarid terrain of north-east Africa. They begin at an elementary level and progressively guide the user through the means whereby TNTlite manages, enhances and analyses digital image data that bear on geological problems. As well as the image data that the exercises use, many of the other digital images on the CD form a resource that instructors can use to create their own exercises and assignments, together with advice on how to import their own data in a TNTlite-compatible form. My choice of TNTlite as a teaching tool was solely because of MicroImages' unique policy of freely distributing fully functional software, and does not constitute an endorsement—readers must make their own judgement.

The CD-ROM includes the full set of stereoscopic pairs of aerial photographs that appear in Chapter 4, in the

form of anaglyphs that can be viewed stereoptically using the viewing spectacles packed with the CD-ROM. This supplements the need to use a lens stereoscope to obtain full benefit from the text figures. Finally there is a collection of various types of image that contain important geological features.

Because there may be readers who wish to study the book simply as a textbook, and some who have no easy access to a computer, the text does not depend on the CD. At the end of each chapter is a brief guide to the relevant CD-ROM resources. Appendix C gives instructions for installing TNTlite, and also appears as Resources.rtf on the CD-ROM.

Remote sensing roughly means extending human sensory perception to distances greater than we can achieve unaided and to information that is far beyond our physiological capabilities. Vision is far and away our most powerful and flexible sense, and so the strict focus is on capturing information about the Earth's properties in the nearly continuous, two-dimensional fashion that is characteristic of images. This is possible only for those properties that control how Earth materials interact with electromagnetic radiation—not only visible light, but a spectrum that extends from gamma- to microwave radiation. Other attributes relating to natural variations in density, magnetic and electrical properties are detectable, but only in a discontinuous fashion—from point to point or along surveyed lines. The same holds for variations in chemistry, topographic elevation and the geometric structure of rocks, both at and beneath the surface, plus a host of other kinds of geological information. Although some of these attributes can be measured from a distance, the immediate results are not images. They are not excluded from the book, however, because there are means of recasting numbers distributed irregularly in two cartographic dimensions into the form of images. Visual perception is unsurpassed in extracting nuances from any kind of image, whatever its means of derivation. So, there is an overlap between remote sensing and more familiar means of gathering geoscientific information. Part of it is through images, and part through data analysis itself. One of the most important new tools is using computers to find patterns and correlations among far more layers of information than the human intellect can grasp. We deal as a matter of routine with spatial and to a lesser extent time dimensions, but a geological problem often involves tens of different dimensions. The vast bulk of the information is distributed in terms of geographical co-ordinates—it can be registered to maps. An extension from remote sensing is a sort of

multidimensional aid to geological skills; a geographical information system (GIS).

Chapter 1 sets out to explain in relatively simple terms how matter and electromagnetic radiation interact. It is these interactions that laid the foundation for designing remote-sensing instruments, planning their use and interpreting the images that they generate. Although computers play an ever increasing role in geoscience, they will never replace the geologist. At the end of the day it is the eye and a mental kit-bag of skills, experience and intuition that generate ideas from perception of the real world. Chapter 2 makes some important points about how our visual system functions, and its power and limitations, both from a physiological standpoint and some of its more psychological aspects. This might seem strange in a geological text, but it is the key to matching the information contained within images to the make-up of the interpreter. The necessary background is completed in Chapter 3, which describes many of the instruments involved in generating images. This chapter also reviews the most important sources of images. The Internet is now the single most important means of discovering what information is available, and Appendix C lists sufficient, reliable URLs for beginning a search—many of the sites include Links to others in the same field, and of course the possible sources evolve continually.

Although remote sensing spans much of the electromagnetic spectrum, before 1970 the only public source of images was from aerial photography. The geological use of aerial photographs since the 1930s has left a priceless heritage of interpretative methods, on which remote sensing builds. Chapter 4 is the core of the book and concentrates on photogeological interpretation. Most of the images there are panchromatic photographs, often combined in stereopairs that can be viewed using a standard lens stereoscope (the CD contains all of these as anaglyphs, which you can view using the spectacles packed with it). Some are images derived from digital remote sensing systems, but are used for their content of information on geologically relevant landforms rather than their spectral features. Part of photogeology traditionally has been orientated to the extraction of geometric information and transfer of interpretations to maps; in other words to photogrammetry. Space prevented a proper treatment of this subject, other than aspects of photogrammetry that need to be carried out in the laboratory or the field, which are covered in Appendix A. There has been little advance in photogrammetry for geologists since the 1960s, and I have no hesitation in referring readers to older and more comprehensive texts on the subject, rather than to presume to improve on our forebears.

Remote sensing dramatically entered the geological scene with the launch of sensors aboard continuously orbiting, unmanned satellites in the 1970s. The technology had been available for 10 or more years, but locked

within the military intelligence community. Some of the first civilian systems that emerged with declassification were orientated to a broad and coarse view of meteorology and the oceans. The most revolutionizing aimed at visible and near-visible infrared parts of the spectrum, where vegetation is readily distinguished, and used a resolution less than 100 m to image small areas of the Earth's surface. The Landsat series, initiated in 1972, aimed at providing strategic economic information on the global cereal crop. This never fully materialized, but as well as many vegetation-orientated users, geologists seized on the new, wide and detailed view of landscape. The potential for complete, near-global cover in a standard form revitalized the ambition to extend geological mapping at scales better than 1 : 250 000 from 25% of the continents to the remainder, by using Landsat images as supersynoptic aerial photographs—they are more than just that.

Because the platforms that carried the sensors were unmanned and in quite high orbits, they had to transmit images in electronic form. Rather than using ordinary television methods, the decision was taken at an early stage to code and transmit the data digitally. In Chapter 5 you will grasp the enormous opportunities for 'tuning' digital images to human vision and extracting hidden information from them. It focuses on the spectral properties of materials in that part of the solar spectrum that the Earth's surface reflects. Appendix B complements that by a brief account of digital means of removing the blemishes and distortions that can affect many kinds of electronically captured data.

Conceptually more difficult to interpret are images that depend on the emission of thermal energy by the surface, and the strange interactions of artificial microwaves used in radar remote sensing. These form the topics of Chapters 6 and 7. Perhaps more so than systems operating at the familiar end of the spectrum, thermal and radar studies are attuned to geological phenomena, but there are relatively few geoscientists who are well versed in their intricacies, partly because most images from them derive from experimental airborne systems. The first decade of the 21st century promises to open new vistas with the deployment of such systems on satellites, and an early grasp of the possibilities will allow readers to take advantage of novel opportunities.

Geoscientists collect all manner of information relating to natural force fields, environmental geochemistry and exploration records, as well as lithological and structural data. A gravity determination, an analysis of soil, a bore-hole log or even a measurement of dip and strike at an outcrop are costly. In past decades, such data were gathered for a single purpose, perhaps an exploration programme, then set aside. Simple graphical expression of these variables rarely exploited their full content, and often treated them in isolation. Transforming spatially

distributed data into image form, applying digital enhancement methods, registering many types of data to a common geographical base and then applying multivariate analysis techniques opens new vistas. Not only can the data be revitalized in later programmes, but GIS techniques permit a kind of information fusion that draws together many lines of evidence in solving problems or generating new ideas. Full and often unsuspected value can be squeezed from the records. Moreover, the fusion also helps bring the combined experience and different skills of several people to bear on problems; not any easy task otherwise, as many a manager will verify. This new and growing approach to information forms the topic of Chapter 8.

No course on geological remote sensing would be complete without an explicit guide to its practical applications. Ideally, the best option would be to use actual case histories, but the problem is to précis the best while extracting useful generalizations within set page limits. In Chapter 9, I have chosen to let the experts speak fully for themselves by referring to important case studies, and concentrate on some general principles of strategy and tactics. This covers several topics: landform studies, geological mapping, hydrocarbon, metal and water exploration, engineering applications and studies of natural hazards. They are set in the framework of conventional field surveys, but show how remote sensing can provide vital clues at critical stages.

Remote sensing aided by digital image processing is aesthetically pleasing, because all of us like colourful pictures. If we did not then it would be a pretty sterile

exercise, for it is stimulation of the visual cortex that draws us to find information and then interpret it. However, such sentiments cut little ice with exploration managers and referees of funding proposals. The key to their hearts is cost-effectiveness. A few figures can work wonders. Field mapping in remote areas costs between US\$100–1000 per square kilometre, with an efficiency of between 1 and 10 square kilometres per day, depending on the level of detail. Preparing geological maps at the reconnaissance level using remote sensing costs from US\$0.7–5 per square kilometre, and can run at efficiencies between 50 and 10 000 square kilometres per day, depending on whether airborne or satellite imagery is used and on the information content of the images. Although I must emphasize that remote sensing is no substitute for hitting rocks with hammers, the synoptic view and the access to invisible and rich spectral attributes of surface materials help orientate field work to the most critical areas. They also permit confident extrapolation from visited areas to more remote terrain that no longer needs direct attention. Information fusion through GIS methods brings every conceivable lever to bear on resolving problems and grasping novel opportunities.

Steve Drury
Cumbria, England

NOTE

Some technical terms and concepts appear in the Glossary.

Acknowledgements

Like many geologists, I stumbled into remote sensing out of necessity, in my case by a need in 1977 for an idea of the regional structure of the Precambrian craton of South India. I had no access to aerial photographs or large-scale maps of the region. Without Landsat-2 an ambitious research programme on crustal evolution would have foundered. At about that time Dave Rothery and Mike Baker began postgraduate research at the Open University, which involved using Landsat data in the Oman and the Andes, but had to use image-processing facilities behind the chain-link fence of a secure military establishment. The late Ian Gass, then Chairman of Earth Sciences at the Open University, saw our plight and, typically, more long-term opportunities. His fight for and winning of funds for image-processing facilities laid the groundwork from which this book emerged. Ideas about how to use digital image processing do not fall from trees. The 'spin' of *Image Interpretation in Geology* emerged from shared mistakes, exchanges of ideas and differences of opinion, first with Dave and Mike and later with Gavin Hunt, Barros Silva, Beto de Souza and Margaret Andrews, all research students at the Open University from the 1970s to now. We have been lucky, sharing excellent facilities and funds to get us into the field and to purchase data. Just as important to me have been opportunities to teach geologists the 'black arts' in a number of less well-favoured countries—Syria, Egypt, Arabia, India and Eritrea—which brought home the importance of clarity and demystifying jargon. That experience depended on being invited by the Food and Agriculture Organization of the UN, The US Geological Survey Mission in Jiddah, Oxfam (UK) and the Eritrean Inter-Agency Consortium to run courses in geological remote sensing, and on the trainees, who hopefully did learn something practical.

A book on image interpretation depends as much on its pictures as on its text. A fair proportion stem from my work, but the majority, including some of the most pleasing colour images I have ever seen, were provided from their collections by many colleagues around the world. In this edition, many of the new images owe a lot to resourceful use of image processing by Beto and Margaret. Staff in the Open University photographic and print workshops printed many of the halftone images, and Andrew Lloyd and John Taylor of Earth Sciences drew several of the line figures. The major addition to the book is the set of exercises on the CD-ROM that use TNTlite. Terry Peterson and Randy Smith of Micro-Images Inc (the producers of TNT software) donated a professional version of TNTmips, helped create the CD

and commented from the software designers' standpoint on first drafts of the Exercises. Margaret was the first to test them as a beginner in remote sensing, and pointed out shortcomings of the early drafts. Most of the exercises use Landsat TM data of the newly independent state of Eritrea, where I have been working for 15 years on a variety of practical and research topics. Some of those data were purchased by Jim Howard (formerly of the Technical Department, Oxfam) and Martine Bilanou (formerly Eritrean field director of the Eritrean Inter-Agency Consortium). The stream-sediment geochemical data of part of Scotland, used in Exercise 11, are from the G-BASE archive of the British Geological Survey, courtesy of its Director and thanks to help from Jane Plant and Phil Green of BGS.

Thanks are due to Ian Francis, Sarah Shannon and Dave Frost at Blackwell Science for their work in producing this edition of the book, and to Harry Langford, Bill Houston and Philip Aslett for their careful checking of the text at various stages in the process.

I am grateful to the following individuals and organizations who contributed or gave permission for the reproduction of figures:

Figure 1.5(a) is reproduced from S. Valley, *Handbook of Geophysics and Space Environments*, 1965, McGraw-Hill Book Company; John Wiley & Sons Inc. (1.6, 1.8, 1.12, 1.13, 2.9e, 3.10, 3.14b and c, 3.22b, 3.25, 5.32, 6.1, 8.4); Gordon and Breach Science Publishers (1.7); Economic Geology Publishing Co., University of Texas (1.9, 1.10, 1.16, 1.17, Plate 3.2); G. Bell and Sons Ltd. (2.3, 2.15); Fig. 2.9(d) reproduced with permission from *Photogrammetric Engineering and Remote Sensing*, 49, 3, Daily, p.352 ©1983, by the American Society for Photogrammetry and Sensing; Aerofilms Ltd (3.3, 3.12, Plate 3.1); W.H. Freeman & Co. ©1978 (3.7); J.P. Ford, Jet Propulsion Laboratory, California Institute of Technology (JPL), Pasadena (3.20, 3.27, 4.16a, 7.19, 7.20); H.C. MacDonald, the University of Arkansas (3.22); Stuart Marsh, Sun Oil, Houston (Plate 3.2); Ronald Blom, JPL, Pasadena (7.18, 7.26, Plates 7.3 & 7.4); NOAA NESDIS, Washington DC (3.29); European Space Agency (3.34a); National Space Development Agency of Japan (3.34b); National Space Science Data Center/World Data Center A for Rockets and Satellites, NASA (7.4–7.17a); Fig. 4.1 reproduced from V.C. Miller, *Photogeology*, 1961, McGraw-Hill Book Company; US Department of Agriculture, Agricultural Stabilization and Conservation Service (4.2, 4.38, 4.44a, 4.46, 4.47b, 4.50); Figs 4.5, 4.16(c), 4.43, 4.48, 4.51 and 4.52 reproduced with permission of the Department of Energy, Mines and Resources,

Canada Copyright Her Majesty the Queen in Right of Canada; Clyde Surveys Ltd. (4.6a, 4.11b, 4.12b, 4.31a, 4.45a); National Cartographic Information Center (NCIC), US Geological Survey (4.6b, 4.7b, 4.11a, 4.26, 4.35c, 4.38a); US Geological Survey (USGS) (4.7a, 4.8, 4.10a, 4.15a, 4.18, 4.26, 4.27, 4.35c, 4.38a, 4.45b, 4.47, 4.49); Fig. 4.10(b) reproduced by permission of the Director of Planning, Department of Planning, West Yorkshire Metropolitan County Council; Figs 4.12 and 4.24(a) reproduced by courtesy of Amoco Inc.; Federal Ministry of Works, Nigeria (4.14b); Department of National Development and Energy, Australia (4.14c); I.G. Gass, Open University (4.15); University of Illinois Aerial Photography and Remote Sensing Committee (4.20); Michael Smallwood, Optronix Inc. (Plate 4.2); Dave Rothery, Open University (Plate 4.3, 4.41b, 5.33); P.W. Francis, Open University (4.31b); National Oceanic and Atmospheric Administration, Washington (4.31c); Institut Géographique National, Paris (4.35); Figs 4.36, 4.38(b), 4.40, 4.41 and 4.42 reproduced by permission of the Dirección General de Fabricaciones Militares of Instituto Geografico Militar, Chile; Pat Chavez, USGS (5.18, 5.20); Gary Raines, USGS (Plate 5.7); Environmental Research Institute of Michigan, Anne Arbor

(Plate 5.10); Ken Watson, USGS, Denver (6.6); Anne B. Kahle, JPL, Pasadena (6.7, 6.11, Plates 6.1 & 6.3); Rupert Haydn (Plate 6.2) Fig. 6.15 reproduced from Kahle & Rowan, *Geology*, 8, 1980 Figs 1, 234–9 by permission of The Geological Society of America; Harold Lang, JPL, Pasadena (Plates 6.5 & 9.6); Intera Technologies Ltd, Calgary (7.5, 7.23); Earl Hajic, University of California, Santa Barbara, and Goodyear Aerospace Corporation (7.22); Gerald Schaber, USGS, Arizona (7.24); Tom Farr, JPL, Pasadena (7.25); American Association of Petroleum Geologists (Plate 7.3); Earth and Planetary Remote-Sensing Laboratory, Washington University, St Louis (Plate 8.2b); Charles Trautwein, EROS Data Center (8.16, 8.17); Stanley Aronoff (8.2, 8.6); Oxford University Press (8.3, 8.4, 8.5); Bramley Murton, Institute of Oceanographic Sciences, UK (9.1); JPL, Pasadena (9.2, 9.3, Plate 9.9d); J. Harris, Canadian Centre for Remote Sensing (Plates 9.3 & 9.4); Director of the British Geological Survey (Plates 8.3 & 9.6, geochemical data in TNTlite Exercise 11); Eosat Inc. (Plate 9.8); Trish Wright, Marconi Research Centre (Plate 9.9a); Falk Amelung, Stanford University (Plate 9.9b); Didier Massonnet, CNES, Toulouse (Plate 9.9c); Adobe Systems Inc. (Plate 2.1).

1 Electromagnetic Radiation and Materials

The way in which electromagnetic radiation (radiation for short) is generated, propagated and modified is the key to designing remote-sensing systems, their deployment and understanding the images that they produce. The physical inseparability of radiation and other forms of energy from matter means that when matter and radiation interact, both are inevitably modified in some way. To account for these transformations fully involves complex mathematics with which most readers will not be familiar. For this reason, Chapter 1 takes a simplified look at the phenomena involved. For those readers with the background and inclination, the Further Reading for Chapter 1 provides deeper theoretical insights.

1.1 The nature of electromagnetic radiation

You do not need to know a great deal about the physics of generation and propagation of radiation to begin interpreting remotely sensed data. Light and all forms of radiation behave both as waves and as particles. They have the form of linked electric and magnetic force fields, transmitted in packets (quanta or photons), which have zero mass at rest. A particle of matter, such as an electron, displays wave behaviour under certain conditions. Pure energy and pure matter are totally abstract concepts. Mass and energy are inseparable and are related through Einstein's famous relationship:

$$E = mc^2$$

where E is energy, m is mass and c is the velocity of radiation in a vacuum. This fundamental relationship has been demonstrated in practice by the net loss of mass

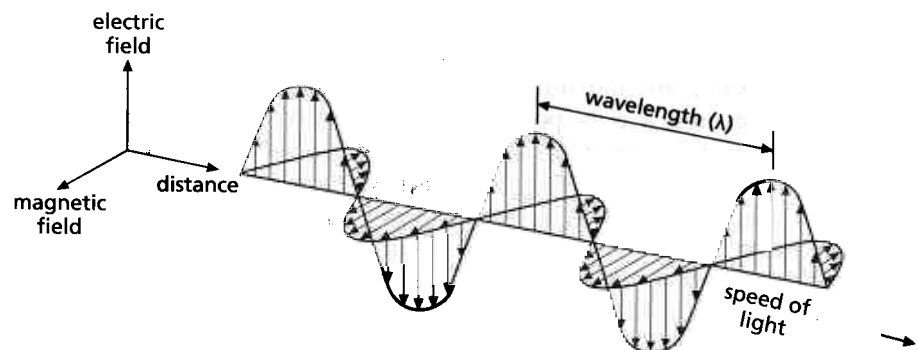
in nuclear fission and fusion, and by the transformation of energy into subatomic particles in high-energy particle accelerators.

More familiar waves, such as sound, ripples on water and those seen moving along a shaken rope, are propagated by the motion of particles of matter. The motion may be back and forth along the travel direction (longitudinal sound waves), in a circulating manner (orbital motion in water waves) or perpendicular to the travel direction (transverse waves in a shaken rope). Those associated with radiation are transverse waves, and involve vibrations at right angles to their direction of travel. However, radiation can travel through a vacuum, and although it can affect particles in a physical medium by changing their electronic, vibrational and rotational properties, it must propagate itself in the absence of matter. Each quantum has associated electric and magnetic fields that oscillate as sine waves at right angles to each other and to the direction of propagation (Fig. 1.1). The distance between wave crests is the wavelength (λ), and the number of vibrations passing a point in 1 s is the frequency (ν). Wavelength is expressed in metres with the prefixes pico (pm = 10^{-12} m), nano (nm = 10^{-9} m), micro ($\mu\text{m} = 10^{-6}$ m) and the rest are familiar. Frequency is expressed as hertz (Hz or per second) with the prefixes mega (MHz = 10^6 Hz) and kilo (kHz = 10^3 Hz). Knowing one enables the other to be calculated from the equation:

$$\lambda\nu = c \quad (1.1)$$

The electric and magnetic vibrations associated with a quantum can be in any orientation at right angles to the direction of propagation. However, if the fields for all quanta are lined up in one direction by some means, the radiation becomes plane-polarized—a familiar concept

Fig. 1.1 Electromagnetic radiation (EMR) comprises waves in electrical and magnetic fields. These fields are at right angles to each other and to the direction of propagation of the waves. The waves represent regular fluctuations in the fields and are described by sine functions. The distance occupied by a complete cycle from peak to peak is the wavelength (λ), and the number of cycles passing a fixed point in a second is the frequency of the radiation (ν). As EMR has a constant velocity in a vacuum (c) it is easy to visualize the relationship between λ , ν and c shown in Equation (1.1).



for any geologist who has used a polarizing microscope for petrography.

The frequency, or wavelength of radiation is a function of the energy of the quanta. Max Planck formulated this relationship (Planck's Law) as:

$$E = \nu h = ch/\lambda \quad (1.2)$$

where h is Planck's constant (6.62×10^{-34} J s). Equation (1.2) means that the shorter the wavelength or the higher the frequency of radiation, the greater the energy of each quantum. As the most important radiation-matter interactions take place at the quantum level, Planck's Law is important in understanding them (Section 1.2).

The final basic property of radiation is its intensity, equivalent to the brightness of visible light. This may be regarded as either the number of quanta or the amplitudes of the electric and magnetic fields. The more quanta at a particular wavelength, the greater the energy that is transmitted. The energy of a single long-wavelength quantum is less than that of one at short wavelength. Consequently, more long-wavelength quanta must fall on a detector to produce a measurable response compared with the number of shorter wavelength quanta that produce the same response. In general, therefore, systems aimed at long-wavelength radiation need to collect radiation either from a larger target area or over a longer time than those used for shorter wavelengths. This has important consequences for the resolution of remote-sensing systems and their ability to discriminate real objects from systematic noise. In reality, things are a good deal more complicated than this, because instruments use different kinds of detectors at different wavelengths.

✱ 1.2 The generation of electromagnetic radiation

As electromagnetic radiation crops up frequently in remote sensing literature, this is a convenient point at which to introduce some of the terminology that is used for radiation. Electromagnetic radiation is a form of energy, and the amount of radiation per unit time (power) is measured in units of J s^{-1} (joules per second), or W (watts). The power incident on or emanating from a body is known as the radiant flux, but it is usually more useful to express it as the power per unit area—the radiant flux density (W m^{-2}). The radiant flux density falling on a surface is known as the irradiance, whereas that leaving a surface is called the emittance (sometimes called the exitance).

Limitations to the size of measuring devices mean that we can rarely measure directly all the radiation leaving a surface. Instead, what is measured is the amount of radiation intercepted by a small detector, which collects radiation travelling through a given solid angle. This

gives the radiant flux per unit solid angle, which is called radiance. The units of radiance are $\text{W m}^{-2} \text{sr}^{-1}$ (watts per square metre per steradian, where steradian is the unit of solid angle).

Sometimes it is useful to consider the quantity of radiation measured at a specific wavelength only. For example the spectral radiant flux is the power received or radiated by a body per unit area per unit wavelength. For convenience, this is often quoted in $\text{W m}^{-2} \mu\text{m}^{-1}$. Similarly spectral radiance is measured in $\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$.

Although it is more correct to quote the amount of radiation coming from a surface as radiance or spectral radiance, when writing or speaking informally or non-numerically the more familiar term brightness is often used instead. This term can refer either to the amount of radiation coming from the surface or to the appearance of a region within an image. For example, if a particular area in an image is said to be brighter than another area, it is clear what we mean, even though we may not be able to quantify the difference in terms of radiance units.

The generation of radiation is essentially a simple process. It is produced whenever the size or direction of an electric or magnetic field fluctuates with time. Radio waves can be produced by the flow of rapidly alternating currents in a conducting body or antenna. The alternation is, in effect, the repeated acceleration and deceleration of electrons. At the shortest wavelengths, gamma-rays result from disruption of atomic nuclei during nuclear fission or fusion reactions. X-rays, ultraviolet radiation and visible radiation are generated by electrons jumping from one stable orbiting shell around an atom to another. When an electron moves from a higher orbit to a lower one, the energy which it loses is converted into a photon of a specific wavelength. Infrared and microwave radiation is produced by thermally induced vibration and rotation of molecules. Microwaves are also generated by fluctuations in electric and magnetic fields.

Wavelengths of electromagnetic radiation span many orders of magnitude, from shorter than 10^{-13} m for the most energetic gamma-rays to longer than 100 km for very long radiowaves. Convenience demands a division of this vast range into several arbitrary regions, each with its own name. Figure 1.2 shows this spectrum of wavelengths and conventional names (with further subdivision of the near-infrared shown later in Fig. 1.5(b)).

In nature all processes that generate radiation are related in some way to the temperature of the body emitting it. All matter in the Universe, even that in the near-perfect vacuum between galaxies, is above absolute zero (-273.15°C) and emits some form of radiation. Just how much is emitted and the range of its wavelengths is a complex function both of temperature and the nature of the body itself. Matter capable of absorbing all electromagnetic energy that it receives and emitting radiation perfectly according to its temperature is known

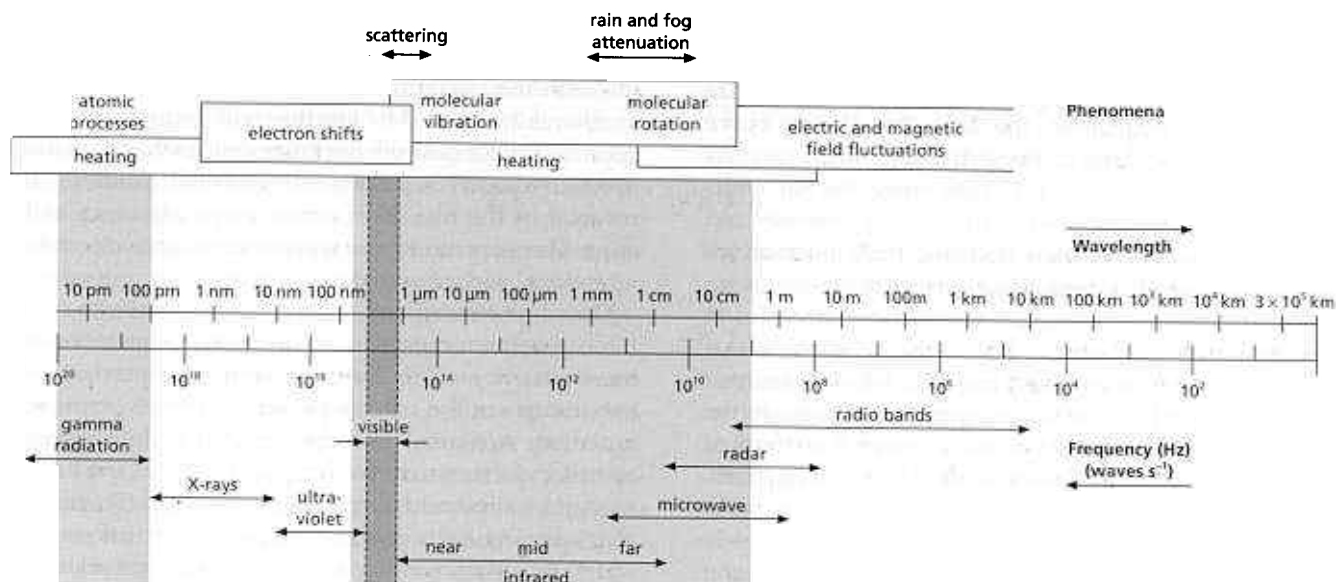


Fig. 1.2 This summary of that part of the electromagnetic (EM) spectrum which is detected routinely by scientists shows the relationship between wavelength and frequency, the phenomena that are involved in generation and interaction of electromagnetic radiation (EMR), and the nomenclature for

different parts of the spectrum (see Fig. 1.5b). Those portions covered by this book are highlighted, together with the processes relevant to geological remote sensing. The narrow visible band is useful as a reference. The wavelength and frequency scales are logarithmic.

as a blackbody. The total energy emitted by a blackbody—its emittance (H) in W m^{-2} —is proportional to the fourth power of its absolute temperature (T in Kelvin or K). This is the Stefan–Boltzmann Law:

$$H = \sigma T^4 \quad (1.3)$$

where σ is the Stefan–Boltzmann constant ($5.7 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$).

At any particular temperature, a blackbody emits radiation with a range of wavelengths. However, its absolute temperature determines which wavelength transmits the maximum amount of energy. This dominant wavelength (λ_m in μm) is given by Wien's Displacement Law:

$$\lambda_m = 2898/T \quad (1.4)$$

So, as temperature increases, total energy emitted rises very rapidly and the wavelength carrying most energy becomes shorter. The shape of the curve relating emittance to wavelength is important (Fig. 1.3), and stems from both the Stefan–Boltzmann and Wien's laws. For any temperature there is a minimum wavelength of radiation, a nearby wavelength of maximum emittance and a long tail towards longer wavelengths. Thus a blackbody at 6000 K—the Sun's surface temperature—does not emit radiation with wavelengths shorter than 0.1 μm , has an energy peak at 0.5 μm (in the part of the spectrum that is visible to us as green), but emits all wavelengths beyond that up to about 100 μm . The total energy emitted is given by the areas beneath the curves in Fig. 1.3.

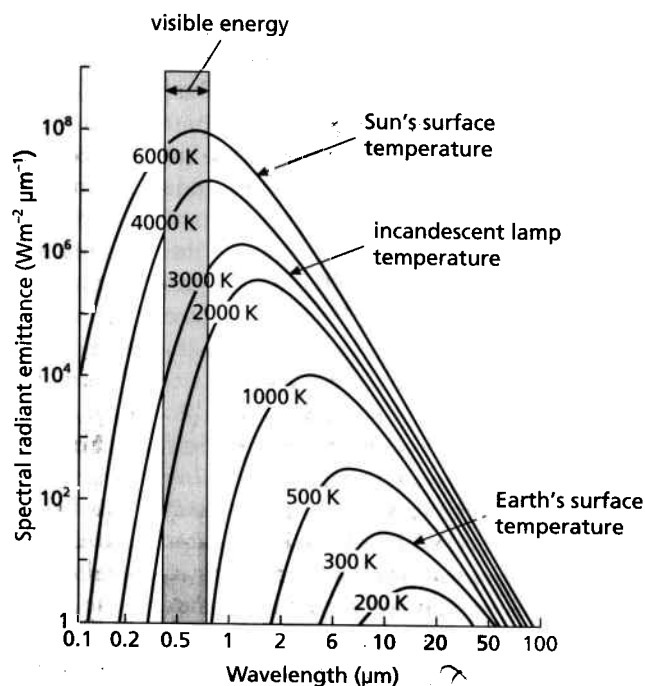


Fig. 1.3 This family of curves on logarithmic axes expresses how energy emitted by a square metre of a blackbody at different temperatures varies with wavelength, and how the wavelength of maximum emittance and the range of wavelengths emitted change with absolute temperature. The area under each curve represents the total energy emitted at each temperature. Both the Stefan–Boltzmann and Wien laws control the shapes.

No natural object is a perfect blackbody. In the case of the Sun—the source of most radiation exploited in remote sensing—many processes other than simple heating are involved. Consequently, the solar irradiation curve (see Fig. 1.5) is a little different from the ideal case. As well as radiation in the 0.1–100 μm range, the Sun emits gamma-rays resulting from thermonuclear processes and long-wavelength radiation resulting from fluctuations in its prodigious magnetic and electric fields.

Remote sensing is concerned with two categories of radiation from the Earth's surface—that which falls on it and is absorbed or reflected, and that which is emitted by the surface itself. Reflected radiation derives in the main from the Sun, and systems detecting it are termed passive because no artificially induced energy is required. An active system involves artificial 'illumination', as in flash photography. In remote sensing the most widely used active systems utilize radar (*radio detection and ranging*) transmissions and detection of the radar energy reflected back to a sensor by the Earth's surface. Experiments have demonstrated that other active systems utilizing artificial radiation, usually in the form of ultraviolet lasers, can produce data for limited ranges of applications.

Because the Earth's ambient temperature is about 300 K, Wien's Displacement Law implies that it has a maximum emittance at 9.7 μm , in the mid-infrared (MIR) range. The energy involved in producing this emitted radiation derives from three sources: the flow of radiogenic heat from the Earth's interior, the heating of the surface by solar radiation, and human activities. Long-wavelength infrared is not the only radiation emitted by the Earth. All rocks, and the materials derived from them, contain variable proportions of the unstable isotopes ^{40}K , ^{232}Th , ^{235}U and ^{238}U , which emit gamma-rays when they decay. These too can be detected remotely, and add to the range of true remote sensing techniques.

1.3 Matter and electromagnetic radiation

The key to understanding how remotely sensed data helps us recognize different materials at the Earth's surface lies in the way radiation interacts with matter. To discuss this exhaustively means examining the interactions at the molecular or atomic level, which involves Erwin Schrödinger's wave mechanics (covered by some of the Further Reading). Only a very simplified account is given here.

For a single chemical element there are several possible states in which it can exist, each of which has a characteristic energy level. Such states involve the types of bonds (covalent or ionic) and the coordination state of atoms within molecules, the energy level of an atom's outermost orbital electrons and much more besides.

The range of states and the associated energy levels are unique for every element and compound. An atom or molecule may pass through a transition from one state to another if it is excited by just the right frequency of radiation. A simple example is fluorescence when radiation of one frequency is absorbed to cause a transition, and reversal of the transition emits lower frequency radiation. There are three basic types of transition—electronic, vibrational and rotational.

Electronic transitions involve shifts of electrons in the outermost orbitals in an atom that impart its element's valency and many of its chemical properties. Such transitions are the reverse of one means of generating radiation. A photon of a specific wavelength induces an outer electron to move from an orbital at the lowest energy level permitted by wave mechanics (its ground state) to one with a higher energy level (its excited state), thereby absorbing the photon. The wavelengths associated with electronic transitions are determined by the principal quantum numbers, angular momentum and spin associated with electron orbitals within a particular element. Electronic transitions occur in solids, liquids and gases, but are particularly important for elements such as iron and chromium, which have several possible valence states, and various positions and coordinations in naturally occurring molecules. These differences account for subtle changes in the wavelengths of electronic transitions depending on the element's host. As electronic transitions require high excitation energies, they are most common at the short wavelengths of ultraviolet and visible light.

Vibrational transitions result in changes in the relative disposition of the component atoms of molecules. The most easy to visualize are distortions of bonds, either stretching or bending from one equilibrium state to another. Analogous to sound, as well as a fundamental wavelength or 'note' associated with a vibrational transition, there are mathematically related harmonics or overtones at other wavelengths. Like electronic transitions, those associated with vibrations of molecular bonds characterize solids, liquids and gases. They require lower energies than electronic transitions, however, and so occur with longer wavelength radiation in the infrared region.

Transitions also may occur in the rotational properties of molecules, but they are restricted to gases. They relate to changes in the moment of inertia of the rotating molecules of gas. Rotational transitions are of most importance, together with vibrational transitions, in the interaction between radiation and the atmospheric gases through which the Earth's surface must be viewed by all remote-sensing systems (Section 1.3.1).

The energy detected by remote-sensing systems over the spectrum of radiation is therefore a function of how energy is partitioned between its source and the materials

with which it interacts on its way to the detector. The energy of any particular wavelength of radiation may be transmitted through the material, absorbed within it, reflected by its surface, scattered by its constituent particles, or reradiated at another wavelength after absorption. In nature, all possibilities combine together in one degree or another.

Three kinds of spectra can be measured for any material—absorption (and its inverse, transmission), reflection and emission spectra. An absorption/transmission spectrum is produced when the material is interposed between source and sensor. A reflection spectrum is measured when both source and sensor are on the same side of the material. For an emission spectrum the material itself is the source. In each set-up a prism or diffraction grating spreads the composite radiation out into its component wavelengths, when intensities at discrete wavelengths, measured by a variety of detectors, may be related to specific emission or absorption processes. This is the technique used by astronomers to detect and measure element abundances in stars from atomic absorption bands in stellar spectra. The remote sensor, however, is more concerned with continuous spectra, which show the variation in energy/intensity over a range of wavelengths. Such spectra are more or less smooth curves in which peaks and troughs indicate maxima and minima around wavelengths that correspond to some characteristic transition. Many macroscopic and microscopic factors conspire together to determine the width, strength and abruptness of these features, some of which are discussed later in this section and in other chapters.

The principle of conservation of energy determines that for any radiation-matter interaction, the incident radiant flux at one wavelength ($E_I)_\lambda$ is distributed between reflection ($E_R)_\lambda$, absorption ($E_A)_\lambda$ and transmission ($E_T)_\lambda$ by the material involved

$$(E_I)_\lambda = (E_R)_\lambda + (E_A)_\lambda + (E_T)_\lambda \quad (1.5)$$

Dividing Equation (1.5) throughout by $(E_I)_\lambda$ produces an expression allowing the spectral properties of the material to be defined in terms of the ratios $(E_R/E_I)_\lambda$, $(E_A/E_I)_\lambda$ and $(E_T/E_I)_\lambda$, which are the spectral reflectance

(ρ_λ) , absorptance (α_λ) and transmittance (τ_λ), respectively, so that:

$$(E_R/E_I)_\lambda + (E_A/E_I)_\lambda + (E_T/E_I)_\lambda = 1 \quad (1.6)$$

i.e.

$$\rho_\lambda + \alpha_\lambda + \tau_\lambda = 1 \quad (1.7)$$

The vast bulk of geological materials are opaque and transmittance is zero, so that Equation (1.7) reduces to:

$$\rho_\lambda + \alpha_\lambda = 1 \quad (1.8)$$

which means in effect that reflectance and absorptance are interchangeable (generally the adjective *spectral* is omitted from both terms), but reflectance spectra are nearly always used. The ratio of the *total* radiant flux reflected by a surface to the *total* radiant flux incident on it (in both cases spanning a range of wavelengths) is known as the albedo of the surface. Although it is not exactly the same, we perceive albedo as the overall visible brightness of a reflective object.

The value of reflectance for a surface specifies the proportion of the incident energy that is reflected at a specified wavelength, but not the direction in which the reflected energy travels. This depends on whether the surface produces specular reflection like a mirror, or diffuse reflection like a sheet of matt paper. In specular reflection, all the reflected energy is directed away at an angle equal and opposite to the angle of incidence. In diffuse reflection, the reflected energy is directed nearly equally in all directions, irrespective of the angle of incidence (Fig. 1.4). A perfect diffuse reflector is often called a Lambertian reflector. Many surfaces combine specular and Lambertian properties, in that they reflect *some* energy in all directions, but reflect a larger proportion in the specular direction (Fig. 1.4c).

A surface behaves as a specular reflector if it is smooth, and as a diffuse reflector if it is rough. Smoothness and roughness depend on the wavelength of radiation. Generally, a surface behaves as rough if its texture is on a scale comparable with, or greater than that of the wavelength of the radiation, and in a smooth fashion if its texture is on a finer scale than the wavelength. Most surfaces, such as rock, soil or grass, are diffuse reflectors in

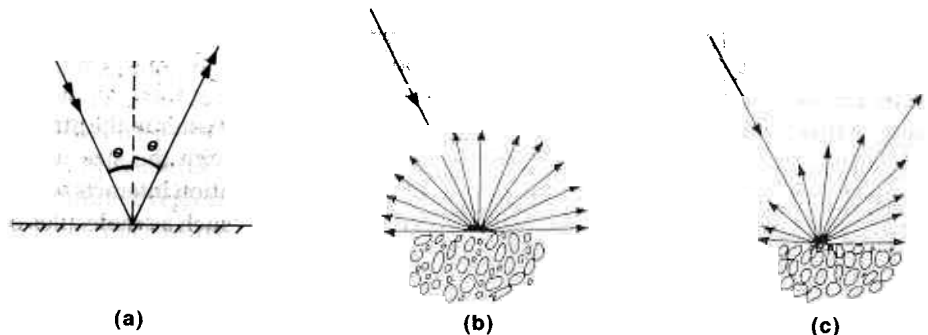


Fig. 1.4 Schematic diagrams showing (a) specular, (b) diffuse or Lambertian and (c) reflection involving both specular and diffuse components.

the visible spectrum—they look equally bright in whatever direction they are viewed, even though tiny parts of a surface (individual mineral crystals, for example) behave specularly.

1.3.1 The effect of the atmosphere

Remote sensing of bodies such as the Jovian moon Io or the planet Mars is a geologist's delight. Both have very thin almost transparent atmospheres, except when volcanic eruptions in the case of Io or sand storms on Mars take place. Virtually the whole of the radiation spectrum is available for surveillance, given a suitable range of sensors. For the Earth, however, all radiation must pass through a dense atmosphere. Before reception by a satellite-mounted sensor, solar radiation must pass down through the atmosphere and then back again to the sensor. For a sensor measuring radiation emitted by the Earth the path is single, but there is some effect, nevertheless.

As well as nitrogen and oxygen, the atmosphere contains significant amounts of water vapour, ozone (O_3), carbon dioxide and traces of other gases. All interact with radiation by vibrational and rotational transitions, the net effect of which is absorption of energy in specific wavebands (Fig. 1.5a). The absorption of short-wave solar radiation is one process that heats the atmosphere. The growth in industrially emitted CO_2 in the atmosphere is the source of the so-called 'greenhouse effect', which is somewhat different. Carbon dioxide's principal effect in this regard is at the longer infrared wavelengths dominated by the Earth's thermally emitted radiation (Fig. 1.3). Outgoing thermal radiation is absorbed by CO_2 and stored temporarily before ultimate re-emission to space. This process is dominated by much more abundant water vapour, but CO_2 is currently increasing so giving rise to fears about global warming. Methane and ozone have a similar effect. This 'lag' in heat loss results in the atmosphere warming above the temperature it would attain without absorbers of thermal radiation.

At short wavelengths atmospheric absorption bands are narrow, but increase in width in the infrared and microwave regions. Figure 1.5(b) shows that about 50% of the radiation spectrum is unusable for remote sensing of the Earth's surface, simply because none of the corresponding energy can penetrate the atmosphere. In the case of emitted gamma-rays, only by flying close to the ground can some energy be detected. It is possible to record absorbed wavelengths relating to gases, but this is for atmospheric studies that are beyond the scope of this book.

Another irritation to the remote sensor is blue sky, strange as it might seem. Obvious when we look upward on a clear day, exactly the same 'tinting' is present looking downwards. It is caused by one of a number of phenomena resulting from the scattering of radiation by

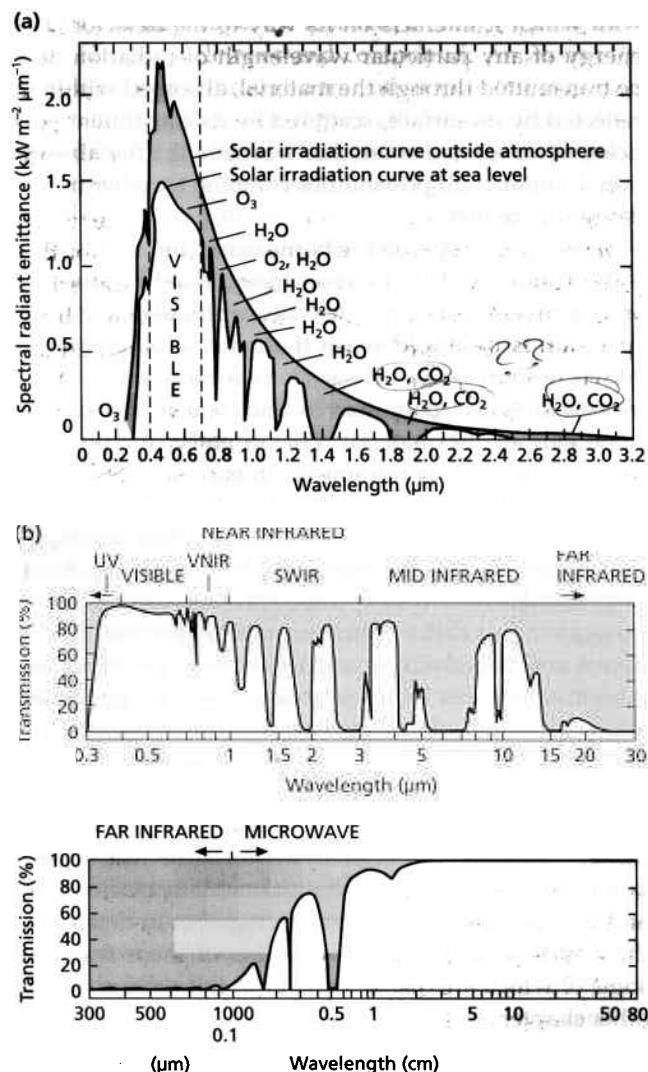


Fig. 1.5 Various gases in the atmosphere absorb solar energy in different wavebands by vibrational and rotational transitions: As a result the solar irradiation curves measured in outer space—upper curve in (a)—and at the surface—lower curve in (a)—are very different. The energy available for interactions with matter at the surface is divided into discrete atmospheric windows separated by bands dominated by atmospheric absorption (grey). In (b) the main atmospheric windows throughout the whole of the useful part of the electromagnetic (EM) spectrum are shown on a logarithmic scale, in terms of the percentage transmitted through the atmosphere. These two graphs, together with the spectral properties of natural materials, form the basis for designing remote-sensing systems.

matter in the atmosphere. The type of scattering changes with the size of the particles responsible. Where radiation interacts with particles smaller than its wavelength, such as molecules of oxygen and nitrogen, the degree of scattering is inversely proportional to the fourth power of the wavelength. This is known as Rayleigh scattering after its discoverer, Lord Rayleigh. The relationship means

that the effects of Rayleigh scattering increase dramatically at short wavelengths—hence blue sky and distant blue mountains. The effect on a view from above the Earth's surface is to swamp the real reflected blue and ultraviolet radiation with a high scattered component, and to reduce the contrast.

Where atmospheric particles are similar in size to the wavelength of incoming radiation, as is the case for the giant molecules of water and for dust, Mie scattering results. This affects wavelengths longer than that of blue light, and is a problem in clear but humid or dusty atmospheric conditions. Red sunsets are attributed to the Mie scattering effect of very fine dust blown up from deserts or microscopic ash particles and acidic water droplets injected into the atmosphere by volcanic eruptions. Aerosol droplets in cloud or fog, which are much larger than most radiation wavelengths of interest in remote sensing, scatter all wavelengths in the visible and infrared spectra. Aerosols are impenetrable, except by radiation beyond 100 μm wavelength—microwave and radar. Even at such long wavelengths, heavy rain or snowfall can cause nonselective scattering of this kind, and as a result can be detected and even measured.

On a clear night stars appear to shift and wink, as do distant objects on a hot day. These distortions are produced by temperature variations in the air, which produce fluctuations in its refractive index, and a range of optical anomalies. These same effects are also present when the Earth is viewed from above. Atmospheric shimmer forms an important constraint on just how small an object remote sensing can detect, irrespective of the theoretical resolving power of each system.

All this implies that remotely sensed images of the Earth are unavoidably degraded in various ways by the atmosphere. Selective atmospheric absorption also means that only some wavebands are available for surveillance (Fig. 1.5b). Those that pass relatively undiminished through the air present atmospheric windows, and determine the framework wherein different methods of remote sensing can be devised. This is an appropriate point at which to introduce a broad division of the spectrum according to the source of the measured radiation in remote sensing. Figure 1.5(a) shows that radiant energy from the Sun falls to very low values beyond about 2.5 μm . Of that in the 0.4–2.5 μm region a good proportion is reflected from the surface, depending on the material, so allowing remote sensing of the properties of the reflected radiation. This is the reflected region. The two windows between 3 and 5 μm and between 8 and 14 μm are dominated by radiant energy that is emitted by surfaces heated by the Sun. This is the emitted region. The transparent region beyond 1 mm is the microwave region. Quite different technologies are needed to acquire remotely sensed data in these three regions, and they are outlined in Chapter 3.

The other constraint on system design, and the most important factor in remote-sensing strategy, is the interaction between radiation and those solids and liquids that constitute the Earth's surface. There are only three important components: water, vegetation and the minerals making up rock and soil. For the geologist, the interactions between radiation and rocks and soils are critical. However, because they can contain water or have vegetation growing upon them, these materials must be considered too. The next three sections introduce some of these surface interactions, and they are expanded upon in later chapters.

1.3.2 Interaction of electromagnetic radiation with rocks and minerals

In this section only the most general features are covered, interactions specific to particular rock and soil types being dealt with in later chapters. Here the effects of common rock- and soil-forming minerals have primary importance. They are considered for three major ranges of radiation wavelength, from 0.4 to 2.5 μm (visible VNIR and SWIR), 8–14 μm (emitted or thermal MIR) and 1 mm to 30 cm (microwaves). Radar interactions in the microwave region are discussed separately in Chapter 7, because they are fundamentally different from those of shorter wavelengths. The most important processes involved are electronic and vibrational transitions (Section 1.3); only gases exhibit rotational transitions.

Rocks are assemblages of minerals, and so their spectra are mixtures of those for each of their constituents proportional to their abundance. Minerals in turn comprise various assemblages of different elements, held together as molecules by different kinds of bonds. Electronic transitions within atoms themselves require more energy than vibrational transitions within molecules, so the former characterize the short wavelength, ultraviolet to visible range, whereas the latter dominate the longer wavelength SWIR. There is, however, some overlap between the ranges of these two fundamental processes.

The most common ingredients of rocks and the minerals that make them up are oxygen, silicon and aluminium, together with varying proportions of iron, magnesium, calcium, sodium and potassium, and smaller amounts of the other elements. Oxygen, silicon and aluminium atoms have electron orbitals in which energy levels are such that transitions between them have little or no effect on the visible to near infrared range. The spectra of minerals are dominated by the effects of less common elements and the molecular structures in which they are bonded. The characteristic energy levels of isolated elements are changed when they are combined in minerals because of the valence states of their ions, the type of bonding and their relationships to other ions

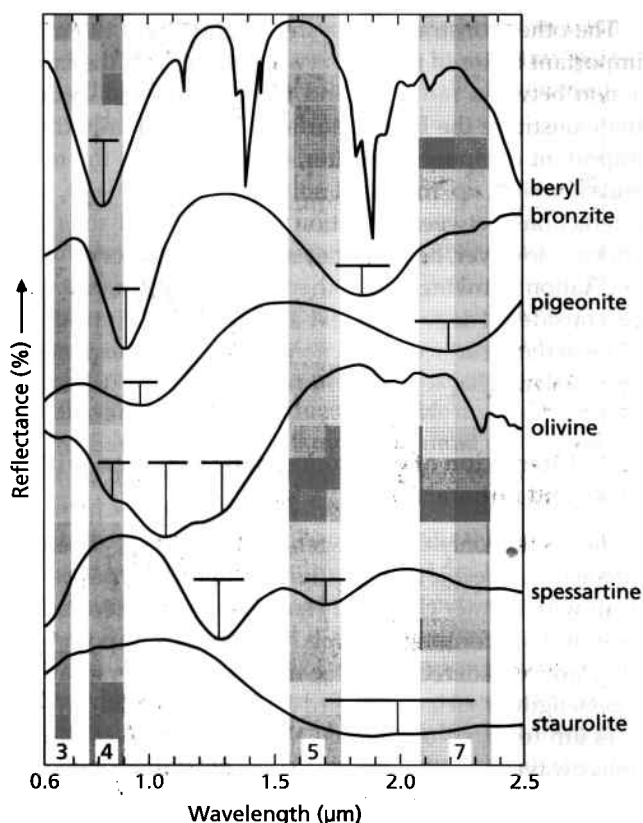


Fig. 1.6 The number and position of features resulting from electronic transitions in iron minerals (T-shaped symbols) depend on the co-ordination of Fe^{2+} ions in the molecular structures of the minerals concerned. The spectra are offset vertically for clarity. In this figure and in Figs 1.7–1.10 and 1.15–1.17 the vertical bands indicate the widths of spectral bands sensed by the Landsat Thematic Mapper (Chapter 3), the most widely used orbiting remote-sensing system.

(their coordination). Because they can exist as ions with several different valences, the transition metals iron, copper, nickel, chromium, cobalt, manganese, vanadium, titanium and scandium exhibit a great range of possibilities. As iron is by far the most abundant of these metals, its effects are the most common and the most obvious.

Figure 1.6 shows reflectance spectra of several iron-bearing minerals that display features resulting from electronic transitions in ferrous (Fe^{2+}) ions. The different wavelengths of the features relate to the symmetry, degree of lattice distortion and coordination of Fe^{2+} in the different minerals. The features are troughs, indicating that absorption of energy takes place over the band of wavelengths involved, in order to shift outer electrons to excited orbitals. All these features are the products of transitions in discrete ions and their breadths result from crystal-field effects. Although there seems to be a wide range of possibilities for distinguishing the iron minerals in Fig. 1.6, in reality they are only useful in a laboratory setting using fresh minerals. In the field, minerals are assembled in different proportions as rocks, so that their

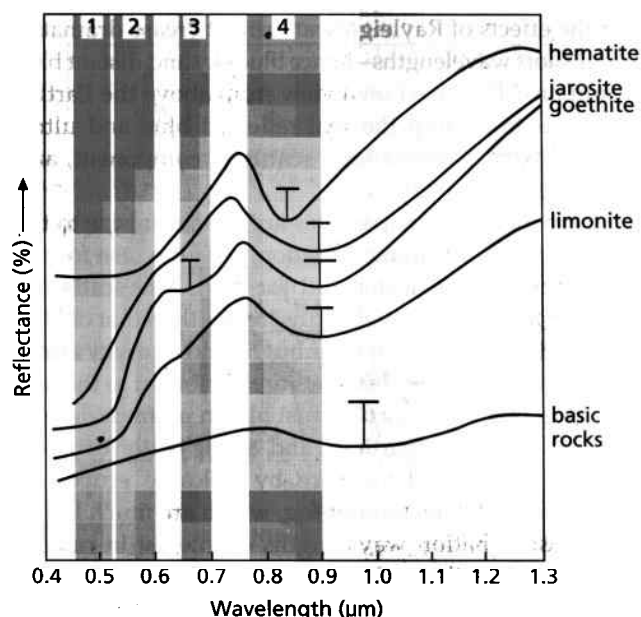


Fig. 1.7 Iron oxides and hydroxides display absorption features resulting from Fe–O charge transfer and crystal-field effects in their reflectance spectra (T-shaped symbols). Substitution of iron in clay minerals superimposes similar features on clay spectra. The spectra are offset for clarity.

spectra interfere. More important, they are rarely fresh but covered with thin veneers of weathering products. As visible and NIR radiation interacts only with the outer few micrometers of the surface, spectra of fresh minerals rarely affect the reflected radiation used in much remote sensing.

Another type of electronic transition results from the presence in metal ions of electrons that have sufficient energy that they are not strongly attached to any particular ion and may transfer from one ion to another. This is the property that imparts high electrical conductivity to metals. In a mineral a similar transition, called a charge-transfer, can occur. It too is induced by energy in narrow wavebands of radiation, giving rise to absorption features. The most common charge-transfer is involved in the migration of electrons from iron to oxygen, and results in a broad absorption band at wavelengths shorter than about $0.55\ \mu\text{m}$. It is common to all iron-bearing minerals and is responsible for a steep decline in reflectance towards the blue end of the spectrum. The most noticeable effect is with iron oxides and hydroxides (Fig. 1.7), and is the reason why these minerals and the rocks containing them are coloured yellow, orange, red and brown. Such minerals form the main colourants in weathered rocks. They too display crystal-field absorptions, the most prominent being around $0.8\text{--}0.9\ \mu\text{m}$. As shown in Fig. 1.7, the location and shape of these bands vary subtly from mineral to mineral and, as discussed in Chapter 5, form a means for discriminating between these important minerals.

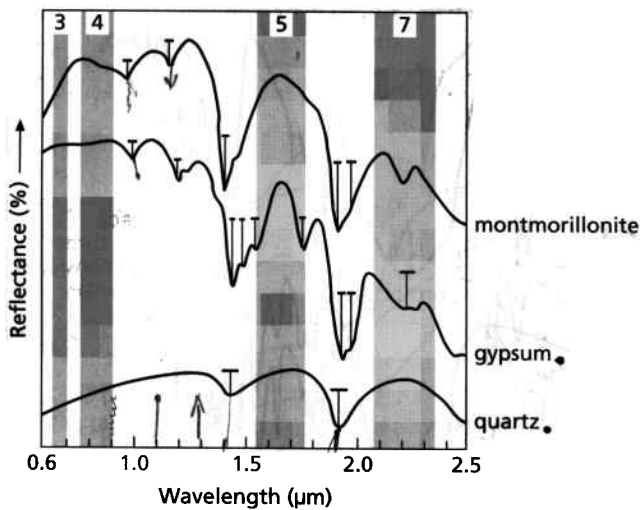


Fig. 1.8 Minerals containing chemically bound waters have particularly distinctive absorption features close to the theoretical overtone wavelengths for the H–O–H bond-stretching transitions (T-shaped symbols). None of the features for sulphates, such as gypsum, relate to SO_4^{2-} ions. The spectra are offset for clarity.

In the SWIR part of the spectrum the most important vibrational transitions in minerals are those associated with the presence of OH^- ions or water molecules bound in the structure or that are present in fluid inclusions. The water molecule has three fundamental vibration transitions, owing to stretching of the H–O–H bond at 3.11 and 2.90 μm and bending at 6.08 μm . As a result of overtones and their combination, these produce absorption features at 1.9, 1.4, 1.14 and 0.94 μm , which are diagnostic of the presence of molecular water in minerals (Fig. 1.8). These features, however, are completely swamped by the nearly identical effects of water vapour in the atmosphere, and are only useful in a laboratory setting.

Many silicates and alteration minerals contain hydroxyl (OH^-) ions in their molecular structure, for which there is only one fundamental bond-stretching transition of the O–H bond, at 2.7 μm . This may form overtones in combination with other transitions, the most important of which are bond-bending transitions, involving distortion of the metal–hydroxyl bonds Mg–OH and Al–OH to produce absorption features near 2.3 and 2.2 μm , respectively. Such features are prominent in aluminous micas and clay minerals (Fig. 1.9) and dominate signatures of hydroxylated minerals that contain magnesium, such as talc, chlorites, serpentines and magnesium-rich clays (saponites). Provided that the absorption features can be resolved, these spectral characteristics form a powerful tool in discriminating between chemically different rock types.

Similar vibrational transitions and overtones characterize carbonate minerals. They derive from stretching and bending of the C–O bond in the CO_3^{2-} ion. They give

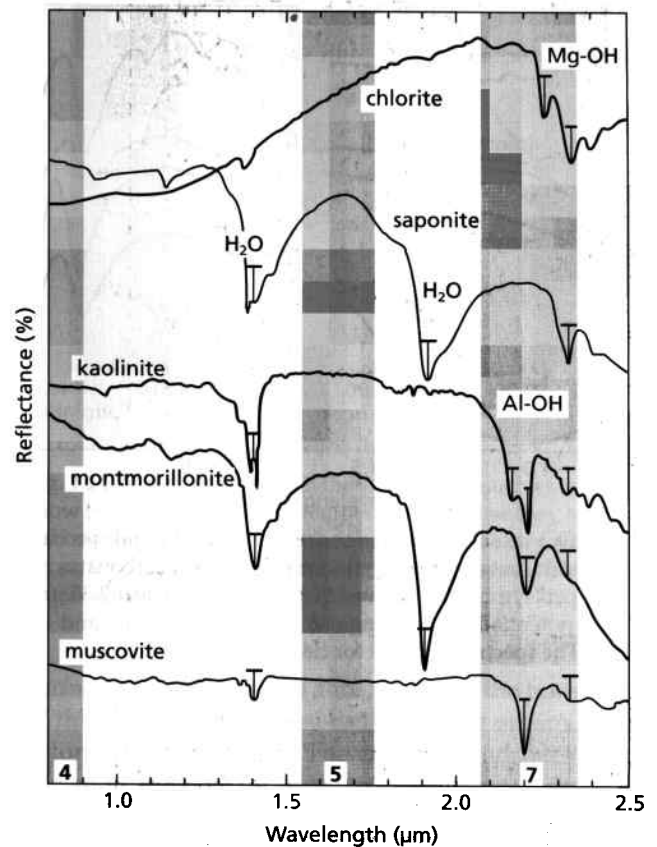


Fig. 1.9 The bending of Al–OH and Mg–OH bonds in clay minerals and micas produce distinctive absorption features in their reflectance spectra (T-shaped symbols). Together with other features of the spectra, they form a potentially powerful means of discriminating these minerals, which are important products of hydrothermal and sedimentary processes. The spectra are offset for clarity.

rise to a number of absorption features in the SWIR of which that around 2.3 μm is most prominent (Fig. 1.10). That at 2.55 μm lies outside the atmospheric window.

Although the reflected part of the spectrum has limited potential for rock discrimination, sufficient diagnostic features are present that great effort has been put into devising methods of capturing data with sufficient spectral resolution to separate the different features. The broad wavebands from the Landsat Thematic Mapper (Chapter 3), highlighted in Figs 1.6–1.10, are capable of detecting but not discriminating between hydroxylated and carbonate minerals. The effectiveness of more sophisticated systems forms an important topic in later chapters.

In that part of the spectrum where the Earth's emitted radiation reaches a peak, an atmospheric window between 8 and 14 μm allows radiation to be sensed remotely. An ideal blackbody, from which the total energy emitted and its distribution between different wavelengths are governed by the Stefan–Boltzmann and Wien laws (Section 1.2), has a distinctly shaped emission spectrum

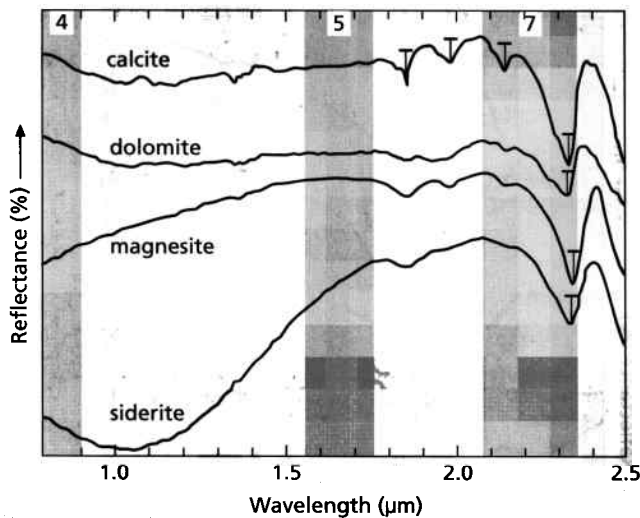


Fig. 1.10 Vibrational transitions related to C–O bonds produce absorption features in the reflectance spectra of carbonates (T-shaped symbols). The most distinctive is that near 2.35 μm , which potentially can distinguish between carbonates and clays. The spectra are offset for clarity.

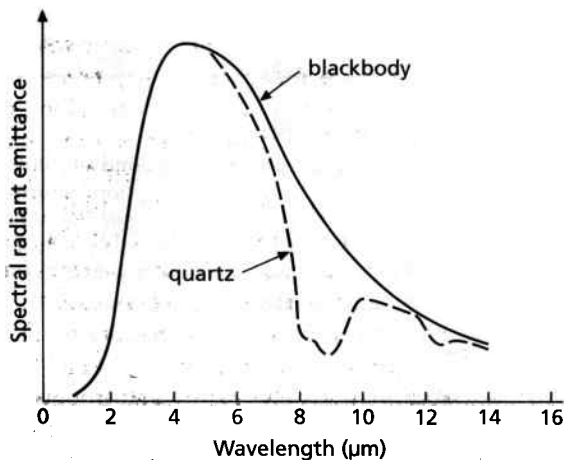


Fig. 1.11 The emission spectrum for quartz, at 600 K, deviates from that for a true blackbody because of a strong feature induced by Si–O bond stretching. Quartz therefore is a selective radiator and not a greybody.

(Fig. 1.3) but no spectral features. A measure of a natural material's deviation from the ideal is its emissivity (e_λ)—the ratio between its radiant emittance for a particular wavelength (λ) at a given temperature and that of a blackbody. A greybody has a constant emissivity less than 1.0 for all wavelengths. Most natural materials, however, have emissivities that vary with wavelength. They are selective radiators because vibrational transitions of bonds in their molecular structure impede emission at characteristic wavelengths. Quartz is a good example, as Fig. 1.11 shows. It radiates as a nearly ideal blackbody up to 6 μm , but deviates from the ideal at longer wavelengths.

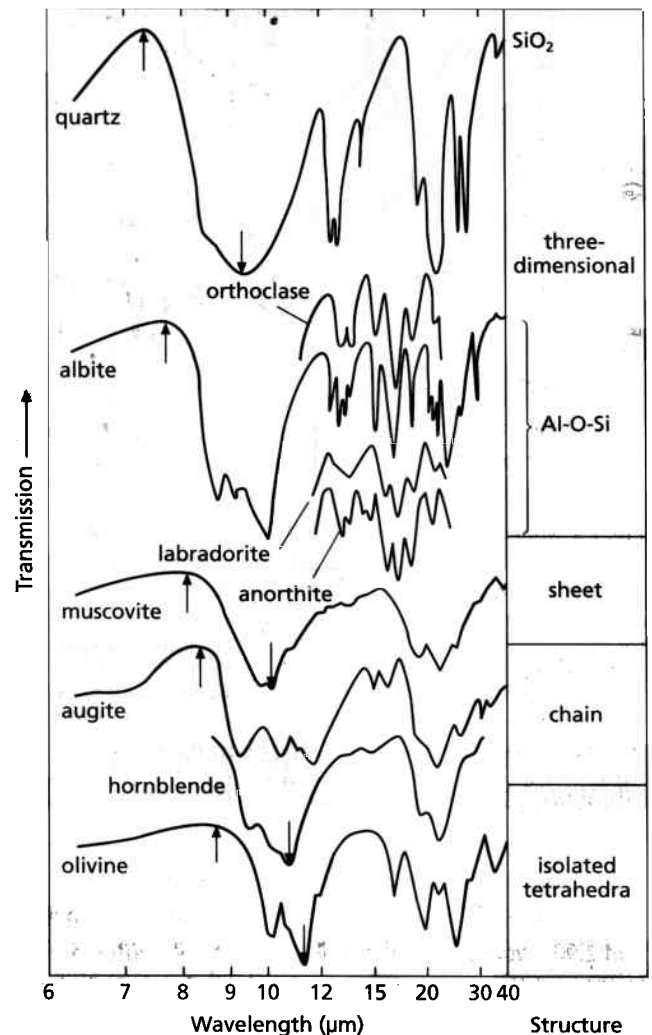


Fig. 1.12 Because of differences in the structure of silicates, the positions of both the Si–O bond-stretching trough and its 'shoulder' at shorter wavelengths (arrows) occur at slightly different positions in the mid-infrared spectra of different silicate minerals. The spectra result from experiments using transmitted energy, but would appear very similar for emission. The spectra are offset for clarity.

Because good emitters are also good absorbers of radiation, their emissivities are equal to their absorptivities. The latter is difficult to measure and this relationship, known as Kirchhoff's Law after its initiator, can be transformed using Equation (1.8) to:

$$e_\lambda = 1 - \rho_\lambda \quad (1.9)$$

The trough in the emittance curve for quartz between 8 and 9 μm is the result of Si–O bond-stretching vibrations. This and related spectral structures are best shown by transmission spectra, and they occur in both silicates (Fig. 1.12) and nonsilicates (Fig. 1.13). This is because most minerals have chemical bonds with energies of vibration that fall in the thermal infrared region, and emission is hindered at the energy which coincides with each bond

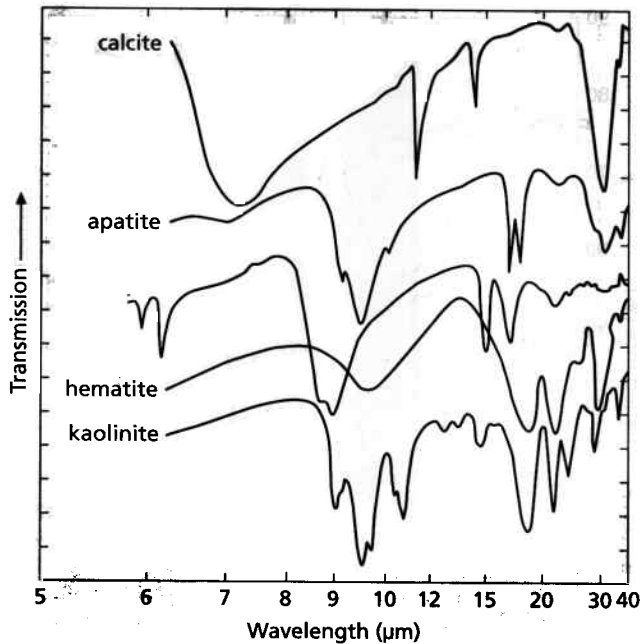


Fig. 1.13 Spectra of nonsilicates in the mid- or thermal infrared part of the spectrum show completely different patterns of absorption features from silicates. They are also very different from one another, suggesting great potential for lithological discrimination by remote sensing of emitted thermal radiation. The spectra are offset for clarity.

vibration. The most important feature in the silicate family of spectra (Fig. 1.12) is that the minimum of the main absorption trough shifts according to the type of silicate structure involved. So too does the peak at the short wavelength edge of the absorption trough. A partial explanation for this is that in silicates with different structures the SiO_4 tetrahedra share oxygens in different ways. The advantage for the geologist is that the progressive shift of the short wavelength peak and the main trough towards longer wavelengths corresponds to a transition from felsic to increasingly mafic minerals (Chapter 6).

In this region of the spectrum, various vibrational transitions in nonsilicates produce spectral features that are different from those of silicates (Fig. 1.13). The most important are those associated with carbonates and iron oxides, which are so distinct that even small amounts of these nonsilicates in dominantly silicate rocks drastically alter their spectra. Indeed, limestones and ironstones should be easily distinguishable from each other, as well as from silicate rocks, in this part of the spectrum.

The total energy emitted by rocks in the thermal infrared region is related to their temperature. By assuming the Stefan-Boltzmann Law it is possible to calculate the radiant temperature of a surface from its total radiant emittance (Chapter 6). The temperature of a rock is contributed to by two sources of energy: the Earth's internal heat flow and solar energy absorbed during daytime. During a 24-h cycle the temperature of the

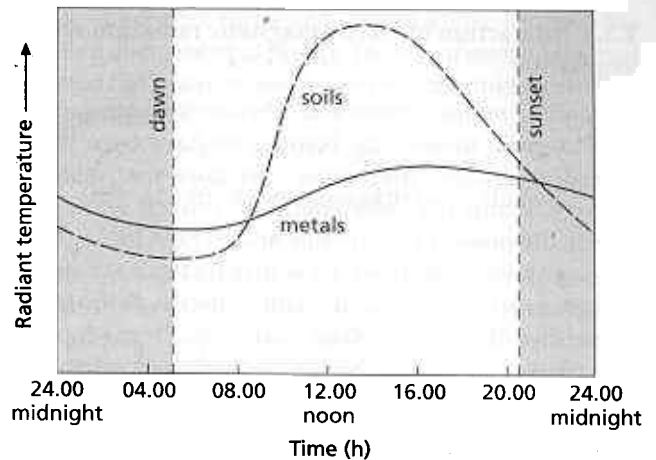


Fig. 1.14 Materials with high thermal inertia, such as metals, show little range in diurnal temperature because they heat and cool slowly. Those with low thermal inertia, such as soils, are prone to rapid heating and cooling, and so they reach high daytime and low night-time temperatures.

surface varies, heating to a maximum at the hottest part of the day and cooling by radiation to a minimum just before dawn (Fig. 1.14). The extremes and rates of this variation depend on a material's absorptance, transmittance and thermal capacity. These variables can be expressed empirically by thermal inertia; a measure of the time-dependent response of the material to temperature changes (Chapter 6). A rock with a high thermal inertia heats up and cools slowly, so showing a low range of diurnal temperatures. Those with low thermal inertia display large fluctuations during the 24-h cycle.

At 300 K the Earth's surface emits radiation at wavelengths in the microwave region as well as infrared, albeit at intensities that decrease to very low levels as wavelength increases. The atmosphere absorbs most of the energy in the thermal range from $14\ \mu\text{m}$ to 1 mm. In the microwave region, however, to which the atmosphere is transparent, intensities are still high enough to be measured by passive remote-sensing systems. A signal detected above the surface will include components emitted by the surface, emitted by the atmosphere and transmitted from below the surface. This last component is possible because rocks and soils have much larger transmittances at these wavelengths than they have in the visible and infrared parts of the spectrum. Thus passive microwave remote sensing is capable of providing information about buried materials as well as surface materials. Its potential use in geology is expanded in Chapter 7.

Rock spectra are composites of those of their constituent minerals. Depending on the structure and composition of these minerals, they may be detected if they are sufficiently abundant and their spectral features are strong enough. Discussion of rock spectra, their interpretation and uses is continued in later chapters.

1.3.3 Interaction of electromagnetic radiation with vegetation

Depending on the climate and whether soils are derived directly from underlying bedrock or have been transported, vegetation may show variations that relate to geology. Plants use solar energy to convert water and carbon dioxide to carbohydrate and oxygen through the process of photosynthesis. How they do this has a strong influence on their interaction with radiation. Being living organisms, their metabolism is strongly dependent on water-based vascular systems and cell structures. The abundance of water in their structure therefore controls these interactions too.

The catalyst for photosynthesis is the pigment chlorophyll, a complex protein containing iron. Chlorophyll absorbs solar radiation to boost the energy levels of electrons and so drive the proton pumping across plant-cell membranes that is the basis for their metabolism. This is achieved by absorption bands near 0.45 and $0.68 \mu\text{m}$ —in the blue and red parts of the visible spectrum (Fig. 1.15). That is why most healthy leaves appear green. In addition to its absorption features, chlorophyll can be made to emit light, or fluoresce, in two narrow bands near 0.69 and $0.74 \mu\text{m}$ if illuminated with a strong beam of light, such as a laser. This is the basis of special laser remote sensing techniques used to assess the chlorophyll content of leaves or plankton. Chlorophyll, however, is unstable above about 70°C . To protect it from thermal breakdown plants have evolved means of balancing energy. This is achieved by strong reflection of near infrared radiation (Fig. 1.15), partly by shiny coatings to leaves but mainly by the internal cells. The structure of plant cells is such that up to 50% of incident, near infrared radiation is reflected internally to re-emerge. The

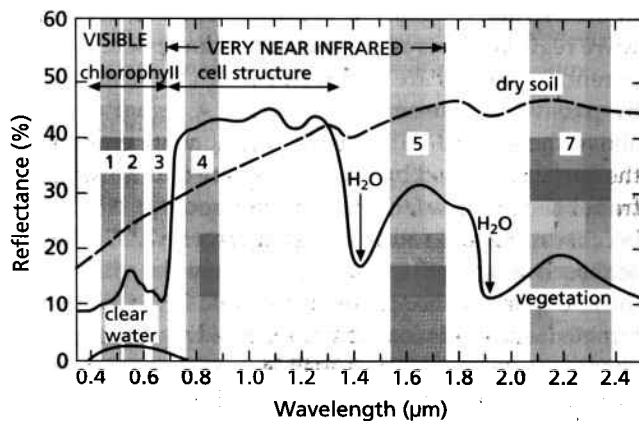


Fig. 1.15 The typical spectral reflectance curve of a leaf shows the strong effect of absorption by chlorophyll in the visible part of the spectrum, the efficiency of reflection of near-infrared wavelengths by its cells and the distinctive absorption features of water contained in its structure. Shown for comparison are spectra typical of soil and clear standing water.

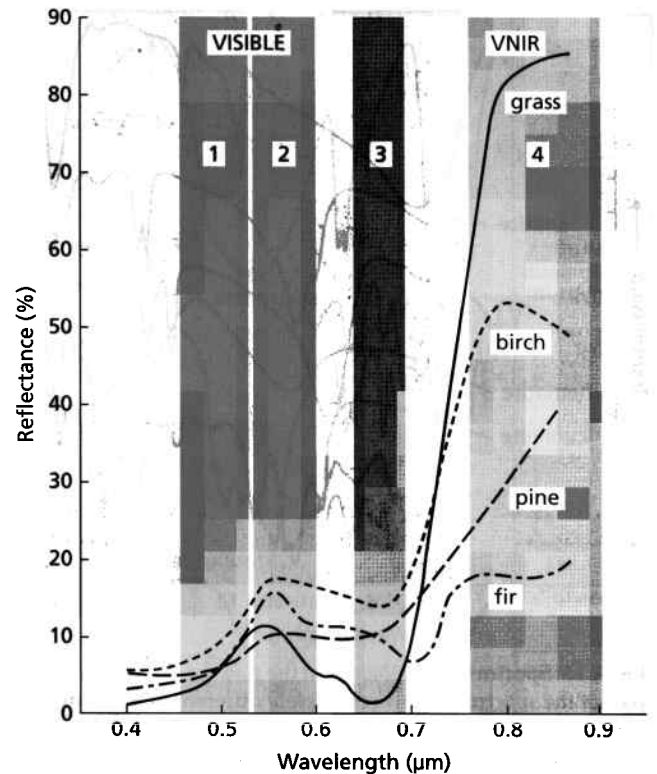


Fig. 1.16 All the different attributes of chlorophyll content, leaf shape, area and number, together with overall plant structure contribute to the spectral reflectance properties of a plant species. Whereas all four plants shown have rather similar properties in the visible spectrum, they are clearly distinguished by their near-infrared reflectance.

remainder is transmitted directly through the leaves. Water in the cells does absorb some energy at its characteristic overtones around 1.4 and $1.9 \mu\text{m}$ (Fig. 1.15), the absorbance depending on the proportion of cell water. Beyond about $2.0 \mu\text{m}$ leaves absorb near infrared radiation.

The different cell structures, proportions of chlorophyll and other pigments, water content and surface morphology of different plants have a marked effect on their spectral properties in the visible to near infrared (VNIR) spectrum. The spectral reflectance of vegetation increases very steeply as wavelength increases beyond about $0.7 \mu\text{m}$ and $0.75 \mu\text{m}$. This sharp change in spectral reflectance is sometimes known as the red edge. Figure 1.16 illustrates this effect for vegetation of different types. It is clear that not only does the height of the VNIR plateau depend on the species of plant responsible, but the exact position of the red edge may vary according to plant type as well. These two factors can also fluctuate within an individual species if the plants are under stress, as a result of deprivation of water or nutrients, or of poisoning by an excess of a toxic trace element such as chromium. Moreover, plants are assemblages of leaves, spaces, twigs and sometimes branches, with different leaf shapes and

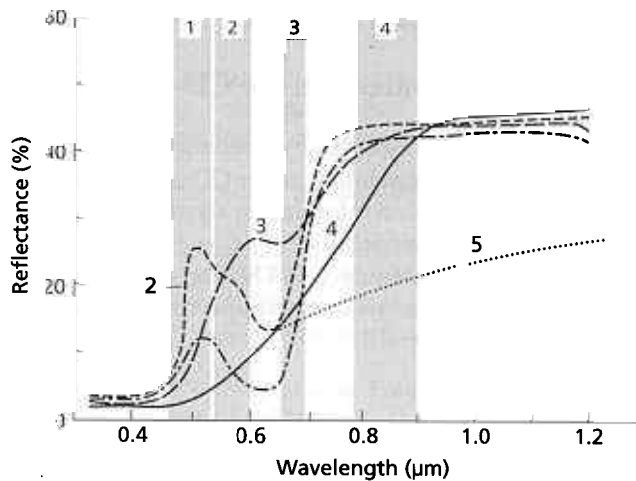


Fig. 1.17 Spectra 1–5 show the progressive stages of colour change in beech leaves preceding autumn leaf fall, from dark green (1) through light yellow-green (2), orange-red (3), brown (4) and dead, dry leaves (5).

sizes and so on. All the individual interactions in such compound structure can interfere and further broaden the range of responses. This facilitates discrimination between species, and between healthy and stunted members of the same species.

As plants have life cycles of various durations their spectral properties are not fixed. As a deciduous leaf matures before falling, its chlorophyll content decays away thereby removing the strong red absorption (Fig. 1.17). Consequently it changes colour from green through yellow to red. As cells shrink and dry they become less efficient at reflecting near infrared radiation. As leaves fall, progressively less of the plant intercepts solar radiation and the reflectance becomes dominated by the soil and leaf litter beneath it. When new leaves appear, near infrared reflectance is well developed, coupled with high yellow reflectance. As chlorophyll content increases the blue and red absorption bands develop, until visible reflectance reaches a minimum at the height of the growing season. Coniferous trees do not shed their leaves during one season, and always display high chlorophyll abundances. As their leaves are small, however, less solar radiation is intercepted and so reflectance of the whole plant is lower compared with deciduous trees in full leaf.

The response of vegetation in the thermal infrared region is complex. Most of the energy absorbed at shorter wavelengths is re-emitted to maintain the energy balance. The radiant temperature of a plant can be as much as 10–15°C above air temperature at night, and up to 5°C below during the day. Plants also control their temperature by transpiration—the exhalation of moisture from pores on their leaves and thereby the loss of latent heat in water vapour. Many factors play a role in determining the rate of transpiration: actual temperature, humidity,

water supply to roots and light (which controls the opening and closing of pores). The use of emitted thermal infrared radiation therefore can provide clues to many of these processes.

1.3.4 Interaction of electromagnetic radiation with water

Bodies of water have a rather different response to radiation than that of water bound up in the molecules of minerals. They do not exhibit the discrete vibrational transition bands so characteristic of molecular water. Instead, the spectral response curves show broad features (Fig. 1.15). In the visible range, the interactions depend on a variety of factors. Just considering reflectance properties, the amount of visible light reflected from a water surface depends on the illumination angle and the presence and nature of waves; smooth water can show 'sun glint' whereas rippled water does so less often. In general, less than 5% of incident visible radiation is reflected by water.

Water has a high transmittance for all visible wavelengths, but it increases as wavelength decreases. As a result, only blue light penetrates water beyond a certain depth, longer wavelengths having been absorbed at shallower levels. In clear water therefore, it is possible to estimate the depth from the intensity of visible radiation reflected from the bottom, particularly that of blue light (Plate 1.1). For depths greater than about 40 m, however, all visible radiation is absorbed and water bodies appear dark.

There is also a certain amount of scattering of light within water, which is responsible for the blue colour of clear water even when it is too deep to see to the bottom. There are two factors contributing to this effect: Rayleigh scattering ensures that short wavelengths are scattered more than longer ones, and the decrease in transmittance with increasing wavelength ensures that scattered blue light is unlikely to be absorbed before it escapes at the surface.

Suspended sediment, plankton and natural dyes, such as tannin from bogs, all increase the reflectance of visible light from water. It is thus possible to estimate the amount of suspended material in water from remotely sensed data.

In the near infrared region water acts almost like a perfect blackbody, and absorbs virtually all incident energy. It is the only natural material with this property, and so water bodies can be distinguished easily from other surface features in this part of the radiation spectrum, even if they are shallow or contain much suspended material. Being a good approximation of a blackbody, water is a nearly perfect emitter of infrared radiation, as well as being a good absorber (Fig. 1.18). This means that measurements of emitted infrared radiation in the

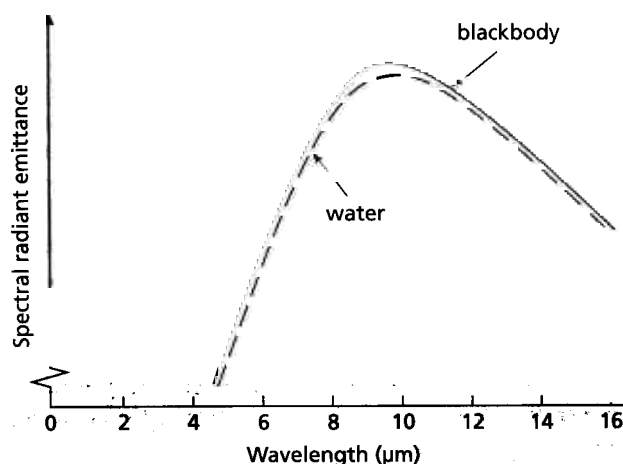


Fig. 1.18 Experimental measurements of the spectral radiant emission of water reveal that it is a near-perfect blackbody. This ought to mean that measurements of energy emitted in the mid-infrared region by water should give a true surface temperature. In practice, however, there are complications owing to the chilling or warming effects of air on the surface film, so that only approximate, relative temperature measurements are possible.

8–14 μm region can be used to calculate the surface temperatures of water bodies very accurately.

A combination of the properties of water in bulk and those of molecular water controls the interaction of radiant energy with water held in the pore spaces of soils and rocks. Pore water increases the absorptance of rock and soil, and so decreases their reflectance. Moist rocks and soils therefore appear darker than when they are dry. In the infrared region, wet soil or rock displays in a muted fashion the reflectance troughs produced by vibrational transitions that are so distinct for minerals containing molecular water and for vegetation (Figs 1.8 and 1.15). In the thermal infrared region, the response of wet soil and rock is complicated by various environmental factors, such as wind cooling and humidity, in an analogous fashion to the effects of transpiration in plants.

Fresh snow is one of the most reflective natural surfaces at visible wavelengths, but in the near infrared region its reflectance decreases with the presence of some of the broad H_2O vibrational features. As snow ages, it recrystallizes, forming larger crystals, and the effect of this is to reduce the reflectance, especially in the infrared region. Ice has a similar spectral response, except that its visible reflectance is below 70%, but if it contains impurities (as in many glaciers) visible reflectance can be below 20%. Changes such as these in the spectral response of snow and ice as they recrystallize with age and incorporate impurities can be readily observed by remote sensing, and are useful in hydrological studies as well as in glaciology. In particular, very old, clean ice is generally blue in colour.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file **Menu.htm** in the **Imint** folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

Exercise 2 uses one of the TNTlite processes that allow you to browse a library of spectral curves for a wide variety of minerals.

Further reading

The lists of recommended reading are not exhaustive, but guide readers towards the most useful and up to date literature relevant to each chapter. Each paper or book contains its own list of references, through which a more in-depth literature survey can be attempted. Examples have been taken from as many different parts of the world and as diverse a range of geological environments as possible. The bias is towards publications that highlight the role of remote sensing in geology. Other application areas, particularly those concerned with vegetation and geomorphology, can often reveal useful new approaches and unsuspected links. Remote sensing is not just a tool that happens to be useful to geologists, but increasingly plays a coordinating role in multidisciplinary studies that unify all aspects of the Earth's natural systems.

General reading

These references comprise the main sources of information on remote sensing in general and a number of geologically orientated texts.

- Beaumont, E.A. & Foster, N.H. (eds) (1992) *Remote Sensing. Treatise of Petroleum Geology Reprint Series No. 19*, American Association of Petroleum Geologists, Tulsa, Oklahoma.
- Colwell, R.N. (ed.) (1960) *Manual of Photographic Interpretation*. American Society of Photogrammetry. Falls Church, Virginia.
- Colwell, R.N. (ed.) (1983) *Manual of Remote Sensing* (2 Volumes), 2nd edn. American Society of Photogrammetry. Falls Church, Virginia.
- Curran, P.J. (1985) *Principles of Remote Sensing*. Longman, London.
- Drury, S.A. (1990) *A Guide to Remote Sensing*. Oxford University Press, Oxford.
- Drury, S.A. & Rothery, D.A. (1989) *Remote Sensing* (Distance-learning course jointly produced by the Dutch and British Open Universities. British Open University code PS670). Open universiteit, Heerlen, The Netherlands.
- Goetz, A.F.H., Rock, B.N. & Rowan, L.C. (1983) Remote sensing for exploration: an overview. *Economic Geology* **78**, 573–590.
- Goetz, A.F.H. & Rowan, L.C. (1981) Geological remote sensing. *Science* **211**, 781–791.

- Gupta, R.P. (1991) *Remote Sensing Geology*. Springer-Verlag, Berlin.
- Henderson, F.B. & Rock, B.N. (eds) (1983) *Frontiers for Geological Remote Sensing from Space*. Report Fourth Geosat Workshop, American Society of Photogrammetry. Falls Church, Virginia.
- Hord, R.M. (1982) *Remote Sensing—Methods and Applications*. Wiley, New York.
- Lillesand, T.M. & Kiefer, R.W. (1987) *Remote Sensing and Image Interpretation*, 2nd edn. Wiley, New York.
- Reeves, R. (ed.) (1975) *Manual of Remote Sensing* (2 Volumes), 1st edn. American Society of Photogrammetry. Falls Church, Virginia.
- Sabins, F.F. (1987) *Remote Sensing: Principles and Interpretation*, 2nd edn. Freeman, San Francisco.
- Short, N.M. (1982) *The Landsat Tutorial Workbook*. NASA Reference Publication 1078. NASA, Washington DC.
- Siegal, B.S. & Gillespie, A.R. (eds) (1980) *Remote Sensing in Geology*. Wiley, New York.
- Smith, J.T. & Anson, A. (eds) (1968) *Manual of Colour Aerial Photography*. American Society of Photogrammetry. Falls Church, Virginia.
- Swain, P.H. & Davis, S.M. (1978) *Remote Sensing: the Quantitative Approach*. McGraw-Hill, New York.

The most fruitful volumes in which to seek literature on every conceivable application of remote sensing in geology are the *Proceedings of the Thematic Conferences on Remote Sensing for Exploration Geology* sponsored by the Environmental Research Institute of Michigan (ERIM). There have been 14 of these conferences up to 2000, and their proceedings contain hundreds of excellent, though unrefereed, papers, which are too numerous to list here. The ERIM also sponsors the annual *Conferences on Remote Sensing of the Environment*, which cover all aspects of remote sensing, including sections on geological applications. Other conference proceedings should be sought by using remote sensing as a keyword in bibliography searches. The main journals for remote sensing are *The International Journal of Remote Sensing*, *Photogrammetric Engineering and Remote Sensing*, *Remote Sensing of Environment* and *IEEE Transactions on Geoscience and Remote Sensing*. Many geological journals contain papers on the results of remote sensing studies. A convenient means of seeking more recent papers on remote sensing is to look for references to older publications using the Science Citation Index.

Further Reading for Chapter 1

These references are the most important for expanding knowledge about the physics of remote sensing, in particular the spectral characteristics of Earth materials.

- Bartholomew, M.J., Kahle, A.B. & Hoover, G. (1989) Infrared spectroscopy (2.3–20 μm) for the geological interpretation of remotely sensed multispectral thermal infrared data. *International Journal of Remote Sensing* **10**, 529–544.
- Hunt, G.R. (1977) Spectral signatures of particulate minerals in the visible and near-infrared. *Geophysics* **42**, 501–513.
- Hunt, G.R. (1979) Near-infrared (1.3–2.4 μm) spectra of alteration minerals: potential for use in remote sensing. *Geophysics* **44**, 1974–1986.
- Hunt, G.R. (1980) Electromagnetic radiation: the communication link in remote sensing. In: *Remote Sensing in Geology* (eds B.S. Siegal & A.R. Gillespie), pp. 91–115. Wiley, New York.
- Hunt, G.R. & Salisbury, J.W. (1970) Visible and near-infrared spectra of minerals and rocks: I silicate minerals. *Modern Geology* **1**, 283–300.
- Hunt, G.R. & Salisbury, J.W. (1970) Visible and near-infrared spectra of minerals and rocks: III oxides and hydroxides. *Modern Geology* **2**, 195–205.
- Hunt, G.R. & Salisbury, J.W. (1971) Visible and near-infrared spectra of minerals and rocks: II carbonates. *Modern Geology* **2**, 23–30.
- Lyon, R.J.P. (1965) Analysis of rocks by spectral infrared emission (18–25 microns). *Economic Geology* **60**, 715–736.
- Lyon, R.J.P. (1972) Infrared spectral emittance in geological mapping: airborne spectrometry data from Pissgah crater, CA. *Science* **175**, 985–988.
- Milton, N.M. (1983) Use of reflectance spectra of native plant species for interpreting multispectral scanner data in the East Tintic Mountains, Utah. *Economic Geology* **78**, 761–769.
- Milton, N.M., Ager, C.M., Collins, W. & Chang, S.H. (1989) Arsenic- and selenium-induced changes in spectral reflectance and morphology of soybean plants. *Remote Sensing of Environment* **30**, 263–269.
- Smith, J.A. (1983) Matter–energy interaction in the optical region. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 61–113. American Society of Photogrammetry. Falls Church, Virginia.
- Suits, G.H. (1983) The nature of electromagnetic radiation. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn pp. 37–60. American Society of Photogrammetry. Falls Church, Virginia.
- Vincent, R.K., Rowan, L.C., Gillespie, R.E. & Knapp, C. (1975) Thermal infrared spectra and chemical analysis of 26 igneous rock samples. *Remote Sensing of Environment* **4**, 199–209.

2 Human Vision

In the field or in the laboratory, geologists rarely use senses other than their sight. Even the surface texture of a rock or mineral is better defined by looking rather than by feeling. With the exception of the distinction between silt and clay by the palatability test (if you nibble a piece of rock with grains too fine to see, the silt is crunchy, whereas clay has the 'mouth-feel' of chocolate!), preliminary classification of rocks and minerals is based on their appearance, either to the naked eye or with the aid of a petrological microscope. To delve much deeper needs purely artificial means—at the simplest by tests of hardness and at the extreme by using a mass spectrometer. Because of this, geologists have had to devise their own highly specialized jargon to organize the visible attributes of rocks. Given an understanding of how natural materials and radiation interact, it is therefore no great hardship for geologists to use their well-practised visual capabilities in interpreting images. Things are that much easier and more efficient, however, if images are produced in a form that is attuned to the characteristics of human vision.

Of all the higher functions of animals, sight is the one with the longest evolutionary history. Although this is not the place to venture deeply into the processes of natural selection that have been involved in the development of vision, the survival advantages of being able to detect and estimate motion, distance and size, and to recognize shapes, patterns and colours are fairly obvious. Eyes have evolved independently in many animal groups, such as the compound eyes of insects and the eerie similarity of those of squids—humble molluscs—to our own. A curious thing about human vision is that although the

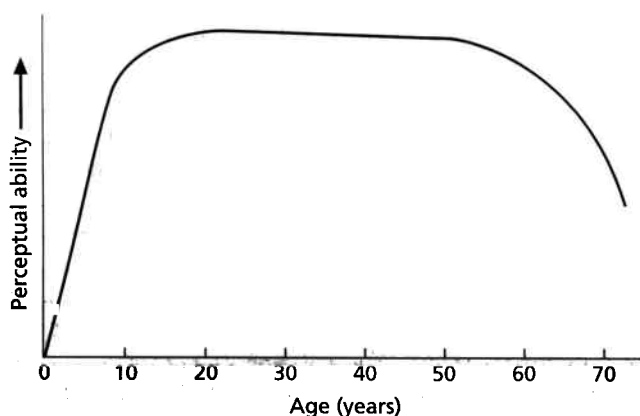


Fig. 2.1 This schematic graph shows the relationship between perceptual ability and age. Most rapid development happens between birth and the age of eight. The ability of adults can be changed significantly by training and new experience.

eye is physiologically capable of superb perception, it has remarkably little innate ability. From the immediate post-natal state of being able to recognize just crude face-shaped patterns, the diversity of human vision, its coordination with other functions and the intellectual capacity to analyse and describe the results are learnt entirely through experience (Fig. 2.1). This means that thoroughly unfamiliar ways of displaying something, like the view vertically downwards from an aircraft, or a representation of the way thermal infrared is emitted by an object, can only be interpreted through tuition, practise and the experience of success. It also means that tonal, textural and shape-related attributes of images, which in the normal course of life would be disregarded, eventually can be extracted efficiently from images, given motivation. In this chapter some of the functions of the human visual system are described in the context of remote sensing.

2.1 The eye and visual cortex

There are few fundamental differences between the optics of an eye and those of a camera (Section 3.1). Both use a lens to form an inverted image on a light-sensitive surface. In the eye this surface—the retina—is a nearly spherical curved surface instead of a plane. The lens in an eye serves merely to adjust focus for objects at different distances and the main refraction is at the cornea (Fig. 2.2).

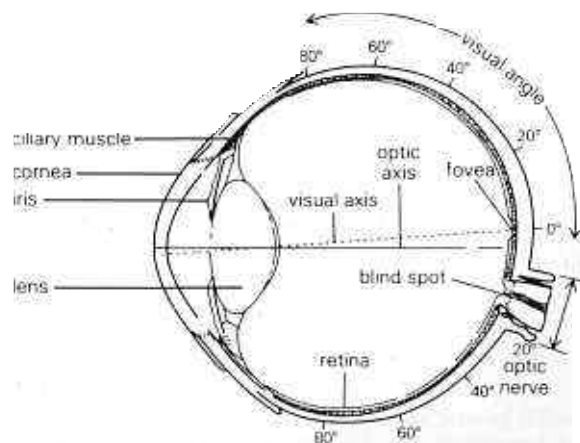


Fig. 2.2 A cross-section of a human eye shows the main elements involved in the focusing of an image on the retina. The chambers in front of and behind the lens contain fluids with different refractive index. They perform most of the refraction in the eye. Muscles, which change the shape and focal length of the lens, are not shown.

The diameter of the retina is about the same as the 24 mm depth of a 35-mm film, so the eye can be compared meaningfully with a 35-mm camera. The lens in a camera that produces an undistorted view has a focal length (f) of 50 mm. That of the eye is about 17 mm, producing a field of view of 180° compared with one of 50° for the camera. The curvature of the retina on which the image is focused compensates for the fish-eye distortion inherent in such wide-angle lenses. The range of distances in focus using a lens—the depth of field—varies inversely with the square of focal length, so the human eye has about nine times the depth of field of a normal camera at the same relative aperture. Rather than the eye elongating and contracting to focus objects at different distances, as happens with a camera, muscles attached to the lens change its thickness and therefore the overall focal length of the eye.

The amount of light gathered by the eye is controlled by much the same mechanism as in a camera. The pupil diameter is changed by muscles in the iris over the range 2–8 mm. These correspond to camera apertures (f -stops) of $f/8$ to $f/2$, respectively. Compare this with the $f/1.4$ to $f/32$ of a sophisticated camera. Pupil diameter also affects depth of field in much the same way as a camera's aperture. It helps overcome optical aberrations, and so improves image definition. Not only the intensity of light governs pupil size, however, but interest and mood too. The optimum pupil size for image interpretation is about 4 mm or $f/4$. There is a great deal more fun in image interpretation than you might think, because this optimum can be achieved readily by occasional visual and, much better, physical contact with an attractive partner.

The retina is the eye's image-recording device, analogous to film and to the charge-coupled device (CCD) arrays in digital cameras. The grains of silver salts or photosensitive dyes in a film emulsion, or individual CCDs are equivalent to some 130 million light-sensitive receptors in the retina (Fig. 2.3). They are not immediately at the surface of the retina. To reach them light has to pass through several layers. The first is formed of optic nerve fibres, which gather to a bundle at the optic nerve (Fig. 2.2). The second is a network of blood vessels and neural cells, which perform some information processing. The receptors themselves are not distributed randomly, nor are they of one single type.

All receptors function through bleaching of a pigment when light is intercepted. The pigment, rhodopsin, has an extremely ancient pedigree, as it is found in purple bacteria that inhabit highly saline lakes and evolved more than 3 billion years ago. The bleaching of rhodopsin reduces the receptor cell's permeability to sodium ions. In the absence of light sodium ions enter the cell to maintain a current, and light decreases the nerve signal. This is the opposite of what we experience, and again seems to reflect the antiquity of vision; for a simple animal,

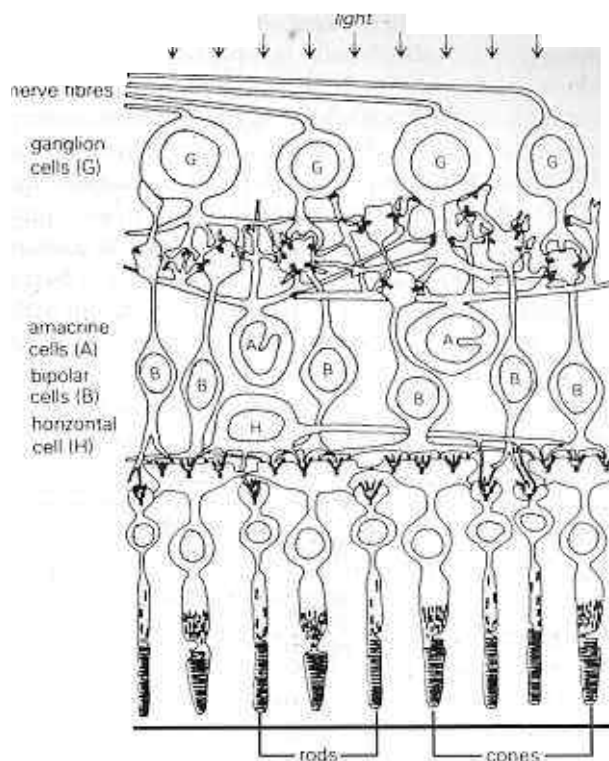


Fig. 2.3 This simplified cross-section of the retina shows the two different types of receptor and a few of the complex neural connections between them.

darkening, as by the shadow of a potential predator, is a more relevant stimulus than lightening. In terms of nerve signals, animal vision is more like a negative than a photographic print.

There are two kinds of receptor. Those equivalent to panchromatic film are known as rods because of their shape. They cover the whole retina, except for the point directly on the visual axis, known as the fovea. Rods are most densely packed at about 20° away from the fovea, and gradually fall off in number towards the edge of the retina. Cones contain three different kinds of pigment sensitive to red, green and blue light. They number only about 7 million, but are concentrated at and immediately around the fovea (Fig. 2.4), although there are some even in the outer area of the retina. At the point where the optic nerve leaves the eye there are neither rods nor cones (Fig. 2.4). This is known as the blind spot. It is rarely noticeable because the brain compensates for the discontinuity.

Rods cannot distinguish colours, and only respond to lightness. The peak of their sensitivity occurs in the blue part of the spectrum (Fig. 2.5), which happens to correspond with the wavelength of maximum atmospheric scattering. Rods therefore are sensitive to the component of surface illumination derived from the sky itself. They are more efficient than cones under conditions of low illumination, when their wavelength sensitivity

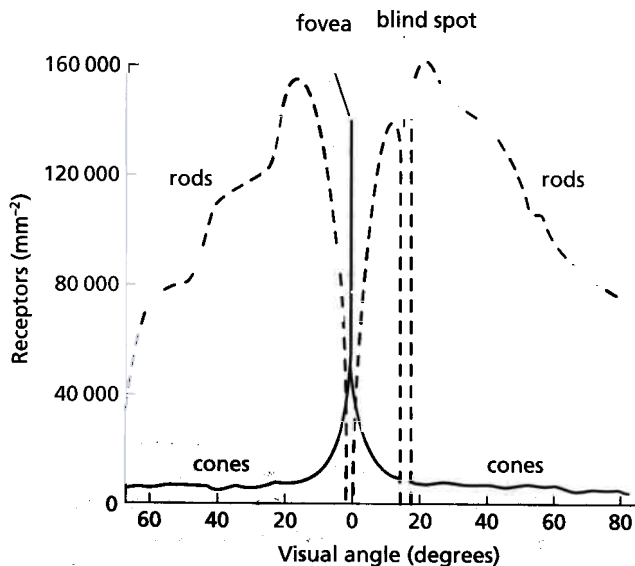


Fig. 2.4 A plot of the density of receptors in the retina shows their uneven distribution, from a maximum around the fovea to a minimum at the periphery of vision.

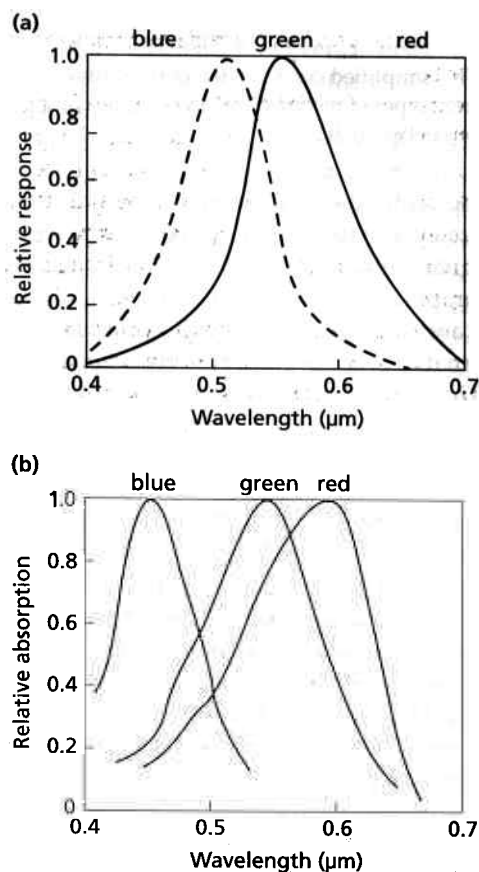


Fig. 2.5 (a) The overall sensitivity of the eye peaks at different wavelengths for scotopic and photopic vision. (b) The spectral absorption curves for the three different kinds of cone, normalized for comparison, show that two peak at blue and green but that normally associated with red light peaks at orange.

comes into its own—at dusk and at night illumination is dominated by sky light. This scotopic or night vision only gives images of brightness, which contain no colour. Blue objects are seen more clearly than red. Definition is also sacrificed in the same way as it is in a fast film. Vision under brighter conditions is described as photopic, and involves both rods and cones. The peak of sensitivity of cones and rods acting together in photopic vision is in the green part of the spectrum and corresponds to the maximum energy region in the solar spectrum (Fig. 1.5). The sensitivity curves for the three main types of cone are shown in Fig. 2.5. Red and green are distinguished by cones only around the fovea, together with blue. Towards the periphery of vision only blue and yellow contribute to colour vision. By having a friend bring a coloured object through your periphery of vision to directly in front of you, while you stare fixedly ahead, you will find that you can see the object in your periphery but only perceive its colour when it is in front of you.

Motion or changes in lightness detected by rods on the periphery of vision signal the viewer to direct the foveal part of the retina towards the source of the change. It then can be examined in colour, with better resolution and with an element of information processing. The neurones, or nerve cells in the retina can transmit impulses directly from the receptors to the visual cortex of the brain, and are in fact an integral part of the brain. Information processing begins on the retina itself. Neurones can inhibit or enhance the signals and also ignore them. In Fig. 2.3 one kind of neurone (H) links many receptors, thus resulting in high sensitivity but low discrimination in the spatial domain or acuity. The same receptors are also connected through more specific neurones (B) for high acuity. Which of the alternative circuits are used by way of a switching mechanism depends on the stimulus itself and return signals from the visual cortex. Some of the connecting neurones respond only to specific shapes, orientations or directions of movement, based on assemblies of receptors forming a receptive field. When they are triggered they cause a specific cortical cell to fire.

The links between receptors, retinal neurones and cortical neural cells are grouped in circuits as cell assemblies. Every new visual experience excites a common core of neurones and in a sense imprints them with the experience. Exposure to a similar experience more easily re-excites the grouping and adds new neurones to broaden the experience. These imprinted neurones are sometimes unconnected with the visual cortex. For instance, exposure to a waterfall in real life imprints an assembly in the cortex based on neurones associated with visual, auditory and even tactile stimuli. Looking at an image of a waterfall later, not only reactivates the visual neurones but can bring back the experience of the sounds and feel of falling water. There is nothing mystical in the feeling induced by a photograph or painting which so closely

represents a waterfall that the observer can almost hear the roar and feel the spray. Producing such a response is an art, but has a base in natural phenomena in the brain as well as in the artist's ability to 'catch' the triggering mechanisms. In turn this is based on the artist's own experiences. In the same way experience of remotely sensed images is built up and art is used in displaying them for best effect and easiest interpretation.

The memory involved is not sited in the brain alone. For simple shapes that build up alphanumeric characters, the basic features of a friend's face or familiar structures in rocks, the neurones on the retina have a memory capacity. Immediate recognition stems from this memory, but more complex analysis is performed in the visual cortex. A complex visual experience is built up in the visual cortex by the response of cortical neurones to particular stimulus patterns on the retina. Different cells respond to different categories of stimulus and each has its own receptive field on the retina. There are three types of receptive field and associated cortical cells. In the simple case a cell responds to a moving stimulus or a spot of light anywhere within the field, but is more responsive to a line or boundary parallel to the particular orientation of the cell's receptive field. Simple cortical cells are line or boundary detectors. Boundaries are the most informative bits of the world, they locate and give outline form to objects. If you were asked to draw a picture of your surroundings, boundaries drawn as lines would be your first step, not building up the scene by complicated colours and shadings. We reduce the world in the manner of a neural conversion to a line-drawing, even though all the nuances of a scene are perceived. That is hardly surprising in an evolutionary sense, for there is little time to ponder on subtleties when a distinctive shape can materialize as a fierce predator!

Complex cells have larger receptive fields and can code directions of lines, boundaries and movements. Hypercomplex cells act as feature detectors for shapes made up of several lines or boundaries, such as the angle between two lines and the angular separation of two lines. They provide the information required in stereoscopic vision (Section 2.4). Acting together, the simple, complex and hypercomplex cells and their receptive fields provide all the information necessary for subconscious and conscious pattern recognition, so important in image interpretation. By the same token they also provide the conditions for optical illusions (Section 2.6), which can have an effect on image interpretation.

2.2 Spatial resolving power

The natural world is a continuum of matter in motion and cannot be broken into its constituent parts in isolation, except by abstract thought. This is expressed by the

particle-wave duality of radiation, Heisenberg's Principle of Uncertainty and several other fundamental relationships. The human visual system produces an imperfect representation of the real world for a whole variety of reasons. Foremost amongst these is the way an image of reality is subdivided into small elements of the natural continuum of spatial and tonal attributes. As the retina is a mosaic of rods and cones, a continuous image falling on to it is recorded in the form of this mosaic. It is dissected into discrete packages. The unconscious scanning action of the eye and the aggregating function of the visual cortex smoothes out the discontinuities, so that the impression is of a smooth whole.

All artificial images are dissected too. They may be made up from brush strokes in a painting, from random grains of metallic silver in a photograph or the dots in the line raster of a television picture. Examining the half-tone and colour images in this book reveals them to be made up of assemblages of ink dots and spaces on the paper. As will be discussed in later chapters, dissection in a regular format is essential for the transmission and computer processing of many kinds of remotely sensed image. Image dissection by the visual system has an important bearing on what the human visual system can and cannot see.

Interpreting the spatial attributes of an image relies on a very complex series of interactions. They involve shapes, patterns, assemblages of tones and colours as different textures, the scale of the image and its context. This function is the most important in image interpretation. In many cases, however, an image contains far more spatial information than the eye really needs in order to make a decision. A newspaper photograph of a pop star or a widely reviled politician contains between 250 000 and 1 million dots, which are sufficient not only for instantaneous recognition but even to make the subject reasonably attractive. On the other hand, a skilled cartoonist or a mimic can reproduce instant recognition by a few sketched lines or suitably exaggerated grimaces. The newspaper image contains much redundant spatial information for the purpose of recognition. However, for the purpose of definite identification—the politician may easily have been substituted by a dummy—the image may not contain enough information. The question, 'Is this the real Adolf Hitler or a rubber doll?', can be resolved only by more information about small physical characteristics, such as the presence of a mole or the precise geometry of a moustache.

For the sake of economy or convenience many images have reduced spatial redundancy, and attenuated tonal or colour redundancy too (Section 2.3). For some interpretative purposes this makes little difference to the observer. Figure 2.6 shows an image taken from an aircraft that has been dissected to give different degrees of spatial redundancy. Some parts of the scene

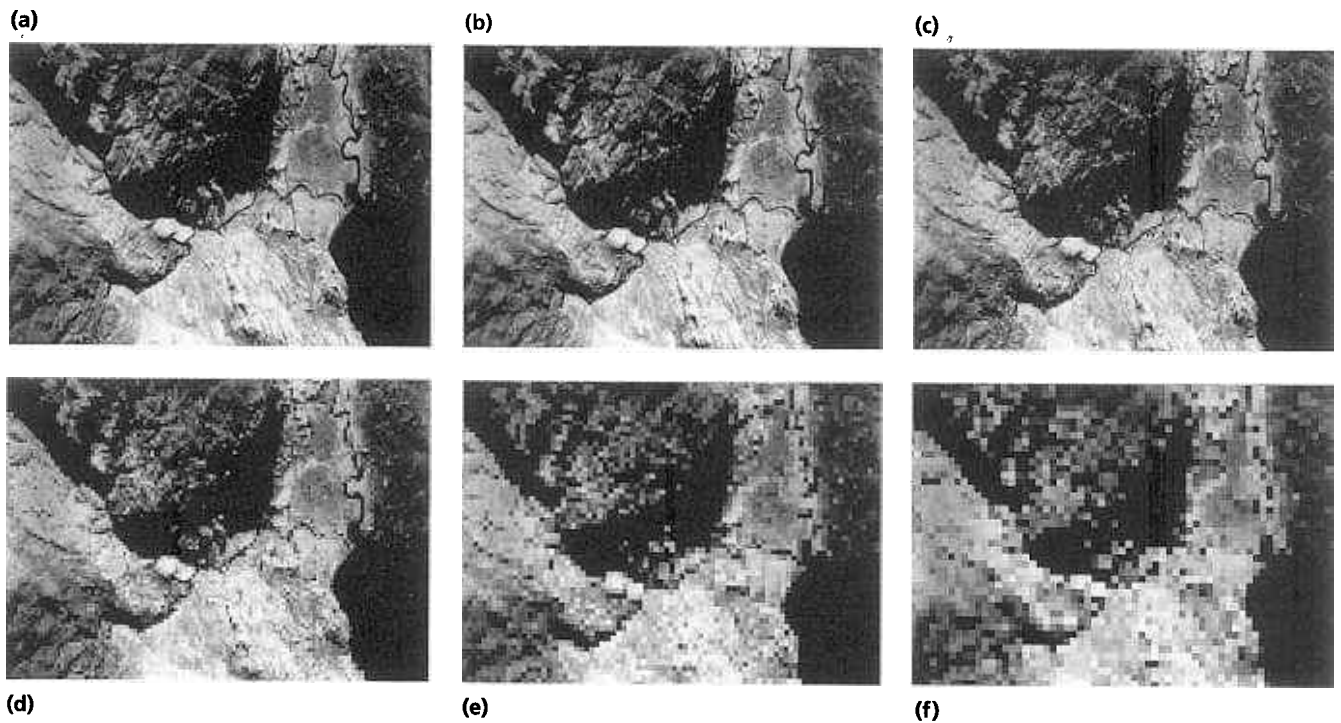


Fig. 2.6 This series of images shows the appearance of the same data after dissection to component picture elements, which increase in size on the ground from 10 m (a) through 20 (b), 30 (c), 40 (d) and 60 (e) to 80 m (f).

are recognizable on all images, others appear only in the least dissected image. Most observers will notice a jump in quality about halfway in the sequence, above which there is no clear improvement, and below which there is little apparent degradation. Viewing the images from increasing distances seems to cause the position of the

transition to change, until all the images appear much the same. On the other hand, increasing the magnification of a dissected image eventually results in the component parts interfering with the impression of the whole (Fig. 2.7). Below this magnification the elements are mentally fused into a coherent whole.

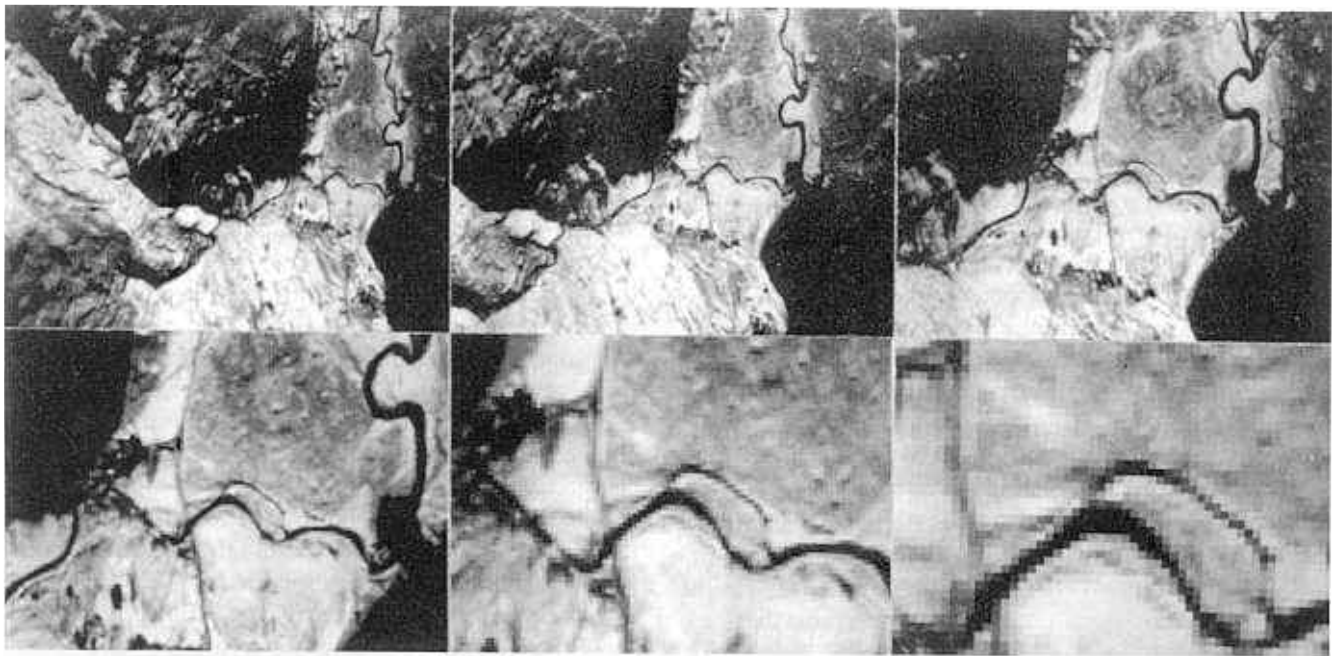


Fig. 2.7 This series of images of the same data at progressively larger magnifications shows the breakdown in perceptual quality associated with the appearance of image dissection.

The chain of processes involved in visual interpretation of the spatial features of an image begins with detection. This is the near-threshold perception that something of interest exists in the image. Being able to define the position of this 'thing' is localization. Recognition that it is a particular category of 'thing', an automobile for example, and identification of its unique attributes—a 1969 VW Beetle with rusted fenders, driven by a friend—depends on observing increasingly intricate features of the object. This may not be possible if the scale of view is too small for the eye to resolve the component parts of the object on the retinal image. At too large a scale the image structure may dominate the view and destroy the visual form of the object. A view of only a few parts of the object in an unfamiliar context may also hinder recognition and identification.

The elements of image interpretation depend on the acuity of the visual system, mental fusion, the degree of dissection of the image, its scale, and on a combination of all these factors, which is expressed as resolution. Another, less well-defined factor is the context of objects within the image.

The acuity of human vision is a measure of ability to distinguish spatial detail. It depends partly on factors inherent in the visual system and partly on purely external factors.

1 Results from experiments depend to some extent on what the subject is called upon to do. Acuity is different for detecting the presence of a test object compared with discerning whether there is one object or two. Likewise, different results occur for localizing the position of the object, and for recognizing what it is.

2 The polarity of the target also has an effect. Bright objects on a dark background are more easily detected than dark on light because scattering in the fluids within the eye broadens a bright stimulus.

3 Contrast is important, a black object on a white background is more easily detected than dark grey on light grey.

4 If the pupil is dilated beyond the optimum of 4 mm, spherical and chromatic aberration in the eye degrades the retinal image. Constriction of the pupil results in degradation by diffraction.

5 The eye takes time to adjust to changed illumination conditions, so there is a time-dependent effect on acuity.

6 The illumination or overall brightness of an image has an effect depending on whether photopic vision is possible or not. Retinal rods have poor acuity compared with cones, and below a certain brightness they take over and scotopic vision is used. Very bright images degrade the acuity of cones as a result of overwhelming the neural chemistry involved in rhodopsin.

7 Under normal illumination conditions the colours in an image help determine acuity. It is higher for greens

and yellows than for reds and blues. This results from the increase in chromatic aberration in the eye at the two extremes of the visible spectrum.

Contrast is defined as the ratio of the difference between the brightness of an object and that of its surroundings to the sum of the two brightnesses, and is expressed as a percentage. The higher the contrast of an object, the more easily it is perceived. Moreover, there is a lower limit of contrast that the eye can deal with, around 0.5%. The effect of contrast is illustrated by Fig. 2.8(a), which illustrates the response of the eye to a point source of energy, such as a star. Although the real boundary between the point source and its background is sharp, the eye's response gradually changes with distance away from the point source, giving a bell-shaped curve termed the point-spread function (PSF). The steeper the gradients involved in this curve, the more likely it is that the point will be perceived. The higher the contrast, the steeper are these gradients and

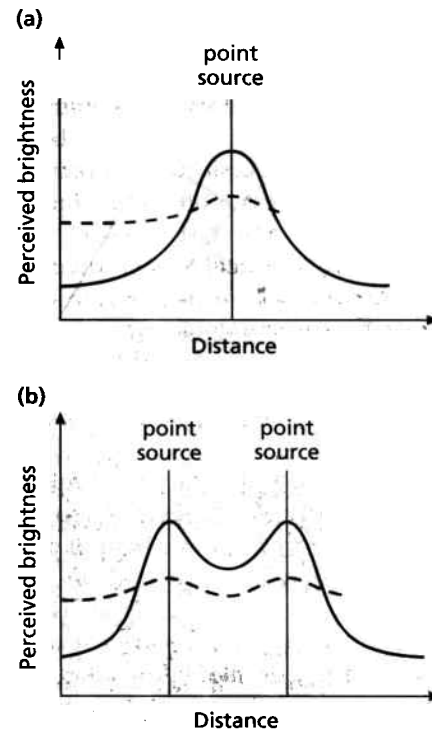


Fig. 2.8 The change in response of an imaging system across a point source of energy produces a variation of image brightness with distance away from the point (a). This is known as the point-spread function (PSF). The more rapid the change away from the point to the level of the background, the better the resolution of the system. The rate of change with distance also depends on the ratio between the energy emitted by the point and by its surroundings. This ratio is the contrast. Low contrast and small brightness gradients (dashed lines) result in poorer resolution than high contrast with high gradients of brightness (solid lines). The two point sources in (b) can be resolved when contrast is high, but are barely distinguishable with low contrast.

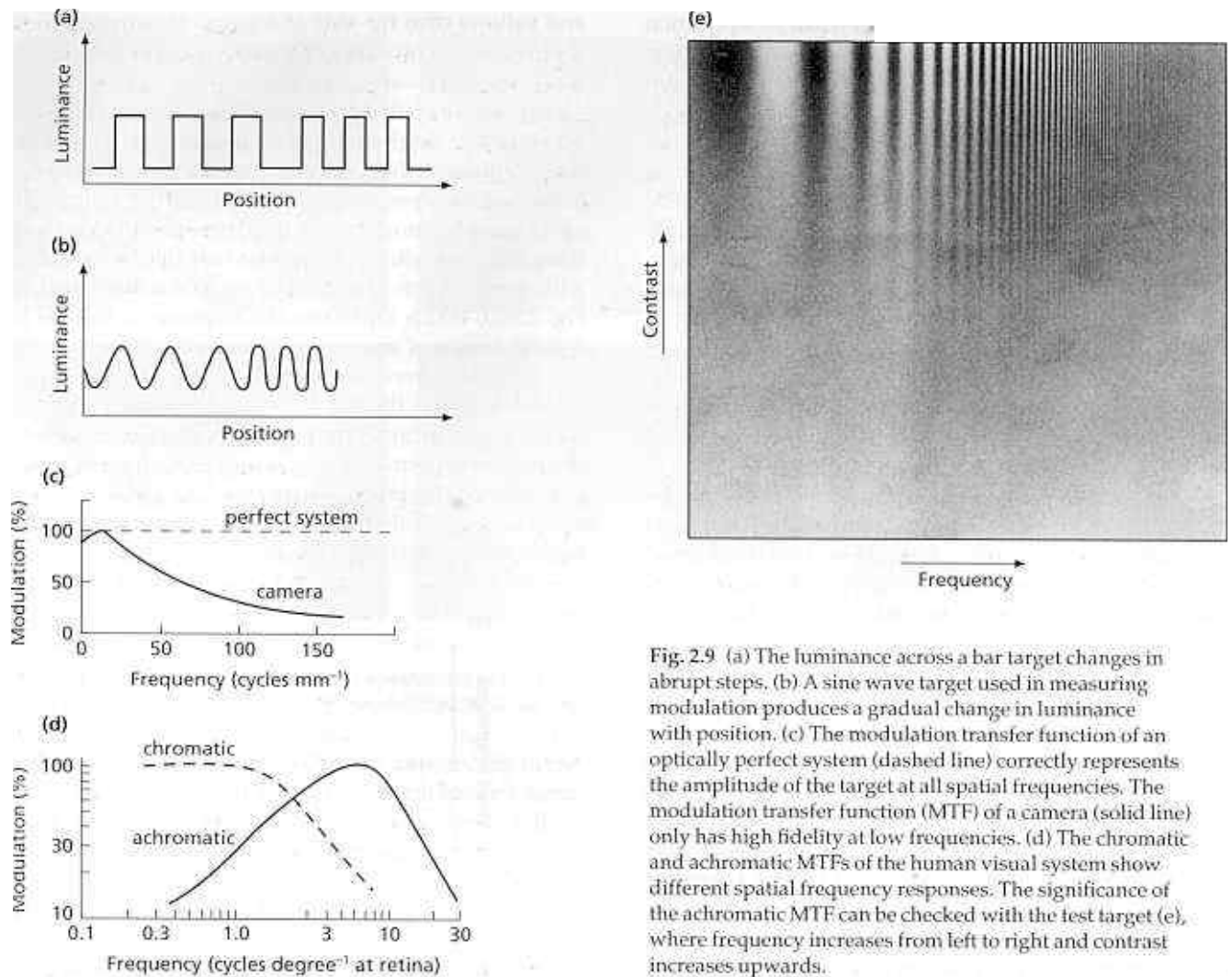


Fig. 2.9 (a) The luminance across a bar target changes in abrupt steps. (b) A sine wave target used in measuring modulation produces a gradual change in luminance with position. (c) The modulation transfer function of an optically perfect system (dashed line) correctly represents the amplitude of the target at all spatial frequencies. The modulation transfer function (MTF) of a camera (solid line) only has high fidelity at low frequencies. (d) The chromatic and achromatic MTFs of the human visual system show different spatial frequency responses. The significance of the achromatic MTF can be checked with the test target (e), where frequency increases from left to right and contrast increases upwards.

vice versa, as shown by the full and broken lines on Fig. 2.8(a), respectively. When the PSF is so smooth that variations are less than the lower limit of contrast perception (0.5%), the point becomes invisible.

Detecting the presence of an object is one thing, but being able to distinguish between several is more important for gathering detailed information. Figure 2.8(b) shows the response of the eye to two point sources. Where each has a high contrast, the response curve (full line) has a distinct saddle between the points, but for low contrast the saddle becomes much less marked (broken line). As the two points become closer the saddle becomes less marked, even at high contrast. The two points become inseparable when the contrast across the saddle falls below the lower limit of perception. So, the eye's PSF and the contrast control our ability to resolve between separate objects.

An interesting series of experiments, which are of direct relevance to image interpretation, measured the response of the human visual system to targets on which

simple spatial variations of brightness were shown. These consisted of regularly spaced spots and lines with different sizes and spacings. Some of the experiments were based on sharply bounded spots and lines, but the most informative approach involved objects with brightness that rose and fell in a sinusoidal fashion (Fig. 2.9b,e). A complete description of the target and the resulting image involves the direction, the wavelength or spatial frequency (the number of waves or cycles per unit distance) and the amplitude (half the difference between the brightnesses at the peak and the trough of a wave). A measure of the accuracy of the detector, in this case the human eye, is the modulation transfer function (MTF).

Modulation is the ratio of the amplitude of the response in the detector to that of the stimulus in the target. The MTF is a graph of modulation against spatial frequency. For a perfect detector the MTF is a horizontal straight line (Fig. 2.9c). A photograph records lower frequencies more faithfully than higher frequencies (Fig. 2.9c), the

actual frequencies depending on the grain size of the film or paper. The MTF of the eye is different for coloured (chromatic) compared with black and white (achromatic) images (Fig. 2.9d). For frequencies lower than 2 cycles per degree at the retina (spacings in the image greater than about 0.2 mm) the chromatic MTF shows that the eye discriminates colour changes in space very well. This colour acuity drops rapidly for more closely spaced objects. The achromatic MTF, however, shows a clear maximum of fidelity at around 7 cycles per degree (spacings in the image between 0.03 and 0.06 mm). A black and white image is obviously better for analysing spatial detail. At higher spatial frequencies the modulation falls because of the physical size and distribution of receptors, and as a result of the blurring effect of unconscious eye movements. The fall at lower frequencies is explained shortly. In a sense, therefore, when looking at black and white scenes, the eye performs as a spatial frequency filter (Chapter 5) and enhances frequencies around 7 cycles per degree at the expense of others. It sorts the trees from the wood, but is unconcerned with the leaves.

In reality, matters are a good deal more complicated and interesting than this. Assemblies of receptors that function to pick up boundaries make comparisons in space between the amounts of light in neighbouring parts of the image. The receptive fields are roughly circular, and therefore operate irrespective of boundary orientation. A central subregion of each responds in the opposite way to an outer 'ring' in the field, thereby enhancing the stimulus. As you will see in Chapter 5, this is how certain types of image-processing filters work with digital images; they were devised before their homologues in the eye were known. Receptive fields vary in size, so that the eye itself performs image enhancement over a range of scales. This goes some way to explaining the difference between the achromatic and chromatic MTFs, rods being more closely spaced than cones and therefore able to form smaller cell assemblies and receptive fields.

Any image or natural scene can be regarded as, and reconstructed from a 'spectrum' of sine waves with different directions, wavelengths and amplitudes. Such a reconstruction is known as a Fourier synthesis, and the disassembly of an image into a family of sine waves is a Fourier analysis (Chapter 5). The spots, lines and bars that form the targets for tests of acuity, and form images too, are themselves made up of sine waves. Small dots and closely spaced lines are dominated by high amplitudes at high frequency. This is why they are difficult to discern. As size increases, high frequency components become lower in amplitude and those in the frequency range to which the eye is attuned increase. Perception and distinction become easier. Beyond this, however, very low frequency components increase in amplitude; effectively degrading perception and distinction.

The eye's MTF is of crucial importance in understanding what can be recognized in an image. It also demonstrates the potential power of artificially enhancing the spatial-frequency distribution of brightness in an image (Chapter 5). Another instructive aspect is that, irrespective of the resolving power of an image-gathering device, the scale at which the image is presented is a major factor in interpretation. The retinal frequency (f cycles per degree) of features on an image is a function of image scale ($1:S$) and viewing distance (X in metres). The following equation relates these to spatial wavelength on the surface (λ in metres):

$$\lambda = SX \tan(1/f) \quad (2.1)$$

The relationship in Equation (2.1), together with the effect of varying contrast, is of crucial importance in designing methods of image interpretation of different terrains. Even the best available image will have little effect on the limitations posed by the human MTF if it is presented at an inappropriate scale. To improve significantly the detail in an interpretation requires a dramatic increase in the scale. To examine very large features means an equally large decrease in scale. For example, a black and white remotely sensed image at a scale of 1 : 500 000 being viewed from a distance of 1 m, gives most visual impact to features with dimensions of about 1.25 km on the ground, i.e. to those appearing on the retina at 7 cycles per degree. Those with dimensions less than 0.5 km and more than 3.5 km will excite less than half the response in the human visual system (Fig. 2.9d). For these purely physiological reasons, therefore, image interpretation is best achieved by combining images obtained from different altitudes, in which the degree of dissection need be sufficient only to avoid the obvious appearance of segmentation at the scale used (Figs 2.6 and 2.7). Using a single image at different scales and viewing distances also helps extract more of the spatial information.

In practice, resolution combines the acuity of the eye with the resolving power of the imaging system, the means of reproduction, the scale of the image and the conditions under which it is viewed. It is defined as the ability to discriminate individuals in a set of small similar items. It is not really measurable in practice, being a threshold between can and cannot. Discrimination of objects is easier using a black and white image than with a colour image at the same scale.

There are some other interesting and useful features of the eye's performance in the spatial domain. First, lines can be separated more easily than points with the same width and spacing. This is because receptors are linked in arrays aimed at line detection. Second, it is easier to detect offsets in lines than to separate lines with a spacing equal to the offset. The absolute limits are 2 s of arc subtended at the eye for offsets and about a

minute of arc for lines. The area on the retina with the greatest power for spatial resolution is near the fovea, where receptors are most densely packed. Optimum interpretation is, as a result, achieved by flicking direct attention from one part of an image to another. In other words, a good interpreter peers at images for long periods, often many times; so they are best pinned up on the office wall!

2.3 Seeing brightness

The eye has a remarkable capacity for detecting light. In a darkened room a flash of light corresponding to 10 quanta absorbed by an area of retina containing about 500 rods can just be detected. This suggests that a single quantum has a realistic probability of exciting a single rod. A rod by itself, however, cannot produce a sensible signal to the brain, but needs the simultaneous response of about 10 other rods. Curiously, this phenomenal detection ability by no means enables the visual system to distinguish brightness variations in a continuous, quantum by quantum fashion.

Experiments on human light detection are based on the noticeability to an observer of changes in the luminance of a lamp or in those of sheets of exposed photographic paper with different grey tones. The term luminance is used to distinguish the objective measurement of the amount of light involved from the perceived brightness of a lamp and the lightness of a reflecting surface. One set of experiments shows that the difference-detection capacity of the eye is nonlinear (Fig. 2.10). The perceived brightness or lightness increases rapidly at low luminance and more slowly at higher luminances.

Another set of experiments focus on the eye's contrast-detection powers. Contrast can have positive or negative values: dark object and light surround, and vice versa. For objects and surroundings with a large visual

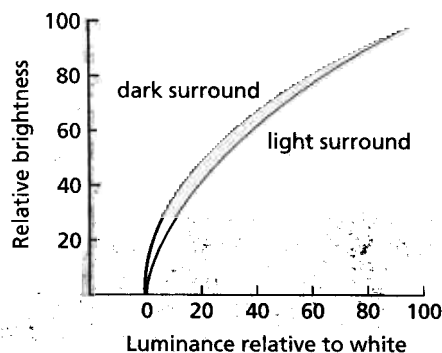


Fig. 2.10 The relationship between the perceived brightness of a lamp and its luminance shows that the difference-detecting ability of the eye is nonlinear.

angle—the ratio between the dimensions of an object and its distance from the eye, expressed in radians—and high luminance, the minimum detectable contrast is 0.5%. This suggests that about 200 steps of contrast can be discriminated in a grey scale made up of sharply bounded luminances. This is very interesting. Although it is a phenomenal detector of light, the eye is not a good discriminator of perceived brightness or lightness. In fact, as the visual angles of objects decrease to those most commonly found in images, the range of discernible contrast differences decreases to about 20–30 steps.

Returning for the moment to objects with large visual angles, it is interesting to note that contrast detection falls off towards lower luminances of object and surrounding. As a result the lightness perceived as the mid-tone of a sequence from black to high luminance in a grey scale is darker than the real mid-tone (Fig. 2.11). The perception of a grey tone also depends on whether the contrast is positive or negative. A grey patch on a black background appears lighter than an identical patch on a white background (Fig. 2.12a). The eye is thus capable of limited contrast enhancement. The reverse effect is produced when fine elements of black or white are introduced into a grey object (Fig. 2.12b). The poor performance of the eye with contrast differences is improved to some extent by a peculiar illusion associated with tonal boundaries. On Fig. 2.11 narrow bands are seen immediately adjacent to the tone boundaries—light bands in the paler blocks and dark bands in the darker. These Mach bands disappear on covering the boundary. Their effect is to enhance and sharpen tonal boundaries. When the boundary is covered it is difficult to distinguish the two adjacent grey tones. The Mach-band illusion stems from the image processing performed by assemblies of receptors, and the same thing happens at tonal boundaries in images when they are filtered by algorithms that are homologous to those in the visual system (Chapter 5).

Black and white photographs or analogue television pictures can contain hundreds or thousands of discrete grey tones. The vast majority are redundant as far as an observer is concerned. If they are progressively reduced, little difference is noticed until the number falls below about 16 (Fig. 2.13). The obvious implication is that the tonal range of a black and white image can be severely truncated, without loss of interpretable information. This is in fact inevitable when high-quality photographs are screened before printing, and when limits have to be imposed on the tonal range of data gathered by digital remote-sensing systems (or in digital television transmission) for the sake of convenience and computing economy. The eye's limitations in the domain of grey tones, however, is amply compensated for by its extraordinary efficiency in the colour domain.

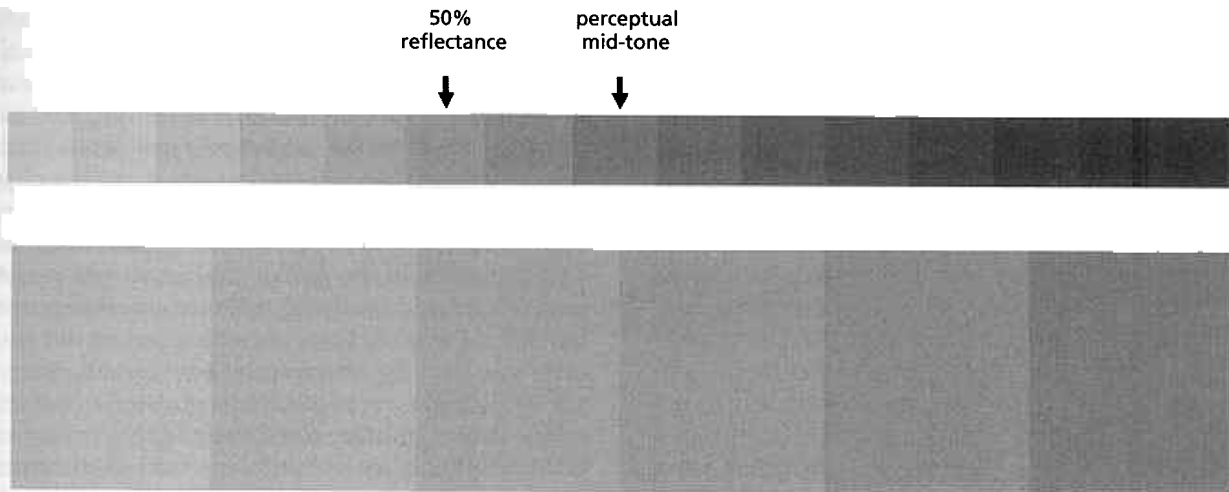


Fig. 2.11 The observed mid-tone between black and white on a photographic grey scale (top) is in fact darker than a grey tone representing 50% reflectance. Another interesting optical enhancement is revealed by the apparent dark and bright lines

slightly to each side of a boundary between grey tones (bottom). These Mach bands disappear when the boundary is covered and it is barely possible to separate the two adjacent grey tones. The steps in reflectance are each 5%.

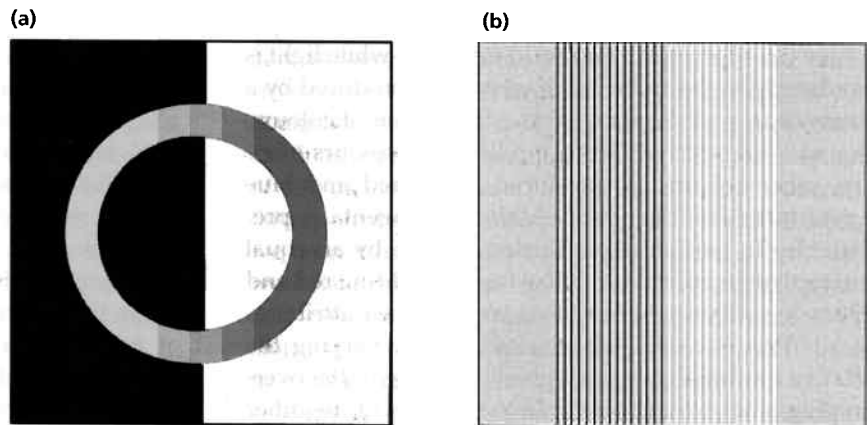


Fig. 2.12 (a) This simple illustration shows the effect of contrast on perceived brightness. The grey circle appears darker against the white background. The effect is more pronounced if the vertical divide is masked with a pencil. **(b)** A uniform grey background appears darker when partly filled with black lines and lighter when white lines are present.

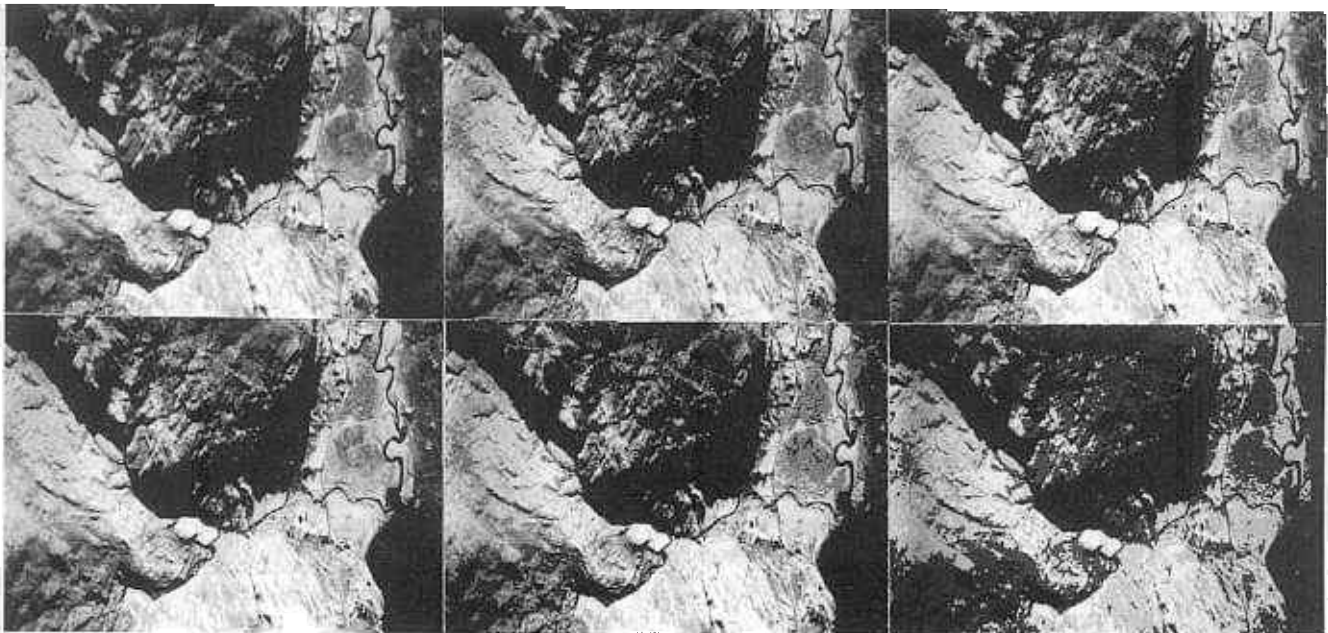


Fig. 2.13 In this series of photographs the same image is shown with 128, 64, 32, 16, 8 and 4 grey levels (from top left to bottom right) to show the futility of trying to perceive more than 20–30 grey tones distributed as small picture elements.

2.4 Producing, seeing and representing colour

Although visual perception is limited in practice to about 20–30 grey tones, the eye can easily distinguish a huge range of different colours. The limit is thought to be about 7 million. Colour dominates human vision. For this reason there is a long history of research into how colour is produced and perceived, enormous difficulty in truly cataloguing colours, and bitter controversy about the contribution of colour to aesthetics. In common practice, it is the source of great annoyance in trying to touch up chipped paint on automobiles—it truly is a law of nature that no autoshop carries matching paint for our cars.

Thomas Young (1773–1829)—famed for Young's Modulus and the wave theory of light—discovered that by projecting light through red, green and blue filters to produce overlapping circles several remarkable effects are produced. First, where all three filters overlap—where the effects of all three add together—white light is produced. As the spectrum of white light produced by a prism or in a rainbow contains a continuum of colours, this was unexpected. Second, where two colours overlap, other colours are produced. For a red and blue mixture the resulting reddish-blue or magenta is predictable. So too is the production of cyan by an equal mixture of green and blue. The result of adding red and green is yellow, which has no red or green attributes at all. This is completely unexpected. By varying the relative luminosities of red, green and blue in the overlapping zone, Young was able to produce all the other spectral colours and many others too, such as brown and pastel shades. These phenomena are illustrated in Plate 2.1. Another of Young's observations was that red, green or blue cannot be produced by mixtures of the other two. Those three colours therefore must be so distributed in the visible spectrum as to constitute separate visual entities. Young termed them additive primary colours, because their effects result from addition. Other combinations of spectral colours also act additively, but only red, green and blue can produce the full range of colour sensations.

A complementary effect can be produced by subtracting red, green and blue components from white light, using filters. Subtraction of red produces cyan, white minus green gives magenta, and the result of subtracting blue produces yellow. All three filters together transmit no light. Magenta plus yellow filters produce red, blue results from equally combined cyan and magenta, whereas yellow plus cyan transmits only green. Mixed combinations of all three subtractive primary colours with different luminances result in the whole gamut of colour sensations.

Additive primary colours are used in colour television screens coated with spots of red-, green- and blue-glowing phosphors. The subtractive primaries are used in colouring paints, in lithographic printing and in reversal colour film. The important point about primary colours though, is not their usefulness nor the physics involved, for the same effects are produced by broad spectral bands in filters as well as by truly monochromatic light. Colour for us is a product of cone receptors in the retina and analysing their combined response in the brain. Young did not know this, but he did realize that there must be a connection between his discovery and the way the eye works. He postulated a tristimulus colour theory, relating the three additive primaries to three different kinds of optical 'nerves', or trichromatic vision. Incidentally, some birds have quadrichromatic vision, using different coloured oil droplets as filters in their cones, rather than having three different varieties of rhodopsin in cones. Our colour vision is in the manner of the goldfish, whereas other animals have no colour vision (achromatic) or vision based on two colours only (dichromats).

That cones have three ranges of sensitivity, peaking in red, green and blue (Fig. 2.14) was not demonstrated until 1964, but this does not explain all perceptual oddities in full. When green is perceived through one eye at the same time as red is seen by the other, the sensation is yellow. This cannot be due to the cones alone. For this reason, and several others, refinement of Young's hypothesis led to the assumption in the 19th century of three pairs of processes involving white-black, blue-yellow and red-green, in the 'opponents' colour theory. This too has been confirmed by experiments, which suggest that

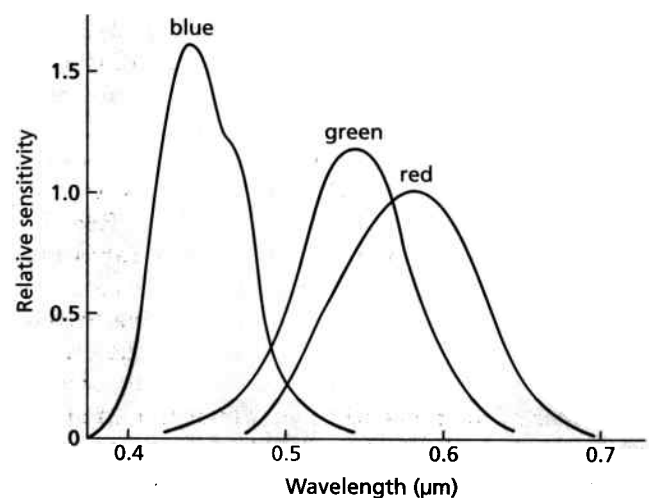


Fig. 2.14 The three types of cone show different spectral sensitivity curves. Although the overall sensitivity for photopic vision peaks in the green, blue cones are most sensitive. This compensates for the smaller number of blue-sensitive cones in the retina.

signals from the cones to the brain take the form of one set encoding brightness signals from all three sets of cones and the rods, together with two sets relating to colour difference. The first of these weights the responses of red and blue cones and encodes their relative strengths. The other is a yellow versus blue signal carrying information from the blue cones and deriving yellow information from the red and green cones. The present consensus is that both the tristimulus and opponents processes are involved. Independent research on the transmission of colour television signals showed that the most convenient method involves a luminance signal and red-green and blue-yellow signals. These are converted in the television to red, green and blue controls of the electron streams emitted by guns focused on the three sets of different phosphors on the screen.

As if this was not sufficiently complex, the visual response to colour has many other peculiarities, some of which are significant to the interpretation of colour images. Some are physiological, others are entirely psychological.

Figure 2.12 showed that perception of brightness is altered by the contrast between an object and its surroundings. Very similar effects occur when identical colour patches are placed on light or dark backgrounds, or on differently coloured backgrounds. Not only are the apparent brightnesses different, the perceived colours are different too. This effect is physiological. A psychological effect also governs how colours are remembered. In general, light colours are remembered as lighter and dark colours as darker. Impure colours, such as brown, reddish-yellow and bluish-green are remembered as being more strongly tinted by the dominant primary, and are interpreted as more red, yellow or green, respectively. Memories of colour images are more vivid and simpler than reality. This is a source of great irritation among remote sensors; the enhanced colour images that they make by various image processing techniques (Chapter 5) never again look as 'good' as when they first appeared on the computer monitor.

Learning by experience and with guidance provides an association of colours with familiar objects, any perceived deviations from this being disturbing. Pale blue skin, purple bananas and green meat are not attractive. The influence of learned associations is so strong that with lists of names of colours printed in the 'wrong' colours it is almost impossible to identify the colours used without some confusion. Remote sensing covers a wide spectral range in which the reflectance or emittance characteristics of familiar materials deviate from those in the visible range (Section 1.3). Images produced by assigning data from nonvisible bands to red, green and blue give a totally 'false' impression. The commonest example is where near-infrared data are assigned to red, red data to green and green to blue in a monitor display. The appearance of vegetation in various shades of red is

aesthetically disturbing. From an objective standpoint it is much more useful than a pleasing natural colour image, however, because of the great variation of infrared reflectance among plants (Section 1.3.3). If the information contained in electromagnetic (EM) radiation outside the visible range is to be exploited, this visual dissonance must be overcome by practice. So, you may ask, why not display the infrared data as green, so that vegetation looks more 'natural'? There is a good reason for not doing so; we are able to perceive a greater range of red shades than green, and blue is the least well perceived of all. Long experience shows that the most visually stimulating colour images use the most information-rich data to control red, and the least to control blue.

Psychophysiology may be the main control over the interpretation of and information extraction from colour images, but it is conditioned by an individual's sense of aesthetics. As a general rule, the most pleasing image is often the most suitable to work with. Human aesthetics is exceedingly curious. Our home planet Earth is regarded as a beautiful and stimulating environment, for the most part at least. The human visual system evolved in this environment. Considered objectively though, it is not particularly colourful. A vast proportion comprises shades of blues, greens, greys and browns. The full capacity of colour vision is reserved for only a tiny volume of flowers, fruits and animals in the natural world. Although we do not have vividly coloured genitalia or plumage, we have developed aesthetics out of natural experience and exercise colour vision much further through art, clothing, furnishings and cosmetics. A few guidelines can be abstracted from this controversial area.

- 1 Many people label some colours as 'warm' and others as 'cool', associated with fire or sunlight, and foliage and water, respectively. There is a general preference for 'warm' colours.

- 2 Given a choice of six spectral colours, experiments show that there is a clear order of preference from most liked to least liked of blue, red, green, violet, orange and yellow.

- 3 For any given colour there seems to be a preference for a lightness that corresponds with the relative lightness of the colour in the spectrum—light yellow is preferred to dark yellow, but dark reds and blues seem more pleasing than light shades.

- 4 How colours are arranged relative to one another in the spectrum appears to govern perceptual harmony. If they are arranged in a circle, adjacent colours, such as yellow and green, or violet and red, harmonize well. So too do diametrically opposite colours, such as yellow and blue. The greatest clash is between combinations of red and yellow-green, for instance, which are about 90° apart.

- 5 Different colours harmonize best when they are equally light and dark, or have the same saturation with a spectral colour.

It cannot be overemphasized that how pleasing a colour image is varies widely from person to person. However, these general rules of thumb, governed by personal foibles, can be used to arrive at the most stimulating and most easily interpreted images of remotely sensed data. Although a certain amount of 'matching' to personality and mood can be achieved during photographic processing, the ultimate control depends on image processing computers (Chapter 5).

The enormous range of humanly sensible colours poses an immediate problem of nomenclature, measurement and cataloguing. Means of satisfying these requirements are necessary for many practical reasons as well as for teaching and research. The major step in systematizing colour was made by Albert Munsell in 1905, when he tried to produce a colour atlas. He based his approach (the Munsell system) on subjective comparison, but it contains the elements of all subsequent colour-coding systems.

Any coloured surface appears to have an attribute related to one of the spectral colours or to nonspectral purple. Munsell designated this the hue of the surface. His system has five principal hues—blue, green, yellow, red and purple or B, G, Y, R and P—and five intermediate hues—BG, GY, YR, RP and PB. These 10 hues are arranged in a colour circle so that the apparent difference in hue between adjacent surfaces are constant. The 10 segments are divided into 10 subdivisions, to give a useful range of 100 hues. A hue coded as 10GY is five steps from green-yellow (5GY) towards pure green (5G) (Fig. 2.15).

Any coloured surface has a luminosity equivalent to a shade of grey between black and white. This is its Munsell value, and ranges from 1 (black) to 10 (white). Colours also appear to be a mixture of a grey and a pure hue. A hue which contains no grey component is said

to be saturated, and it possesses a high chroma. One with a high grey component has a low chroma. The chroma scale in the Munsell system ranges from 0 for achromatic greys to 16 or 17 for the highest conceivable saturation. As Munsell value decreases, so the number of discernible chromas decreases. The Munsell system is therefore three dimensional, as shown in Fig. 2.15, but occupies an irregular solid. This is because it is based on human perception of colour, which is not regular. However imperfect it might be it is still the only internationally accepted standard for visual comparison. It is used by soil scientists, textile and paint manufacturers, and many other groups.

In 1931 the Commission Internationale de l'Éclairage (CIE) produced the nearest approach yet to a fully objective means of describing and measuring colour. The CIE colour co-ordinate system is based on a tristimulus set of co-ordinates, where the three primary colours are imaginary, yet produce a colour space which contains all perceptible colours. The CIE primaries stem from the real primaries red, green and blue. In the same way as all spectral colours can be made from different proportions of monochromatic red, green and blue light, it is possible to express the CIE tristimulus values involved in producing any monochromatic colour by three interfering curves shown in Fig. 2.16(a). For light with equal luminance at any visible wavelength the contributing tristimulus values can be read from the graph. These are divided by their sum to give three chromaticity co-ordinates, x , y and z . These co-ordinates sum to 1.0, so any colour can be specified by just two co-ordinates and a value for the luminance. On a plot of x against y all perceivable colours fall within the curved line shown on Fig. 2.16(b). The numbered section of the line refers to saturated monochromatic light from violet (0.38 μm) to red (0.65 μm). The straight base of the enclosed area represents nonspectral purples.

There are two difficulties with the CIE chromaticity diagram. First, equally perceived differences in colour are not represented linearly. Second, it represents idealized conditions. How the chromaticity co-ordinates of samples are measured to classify materials in this system are beyond the scope of the book.

Various attempts have been made to simplify the Munsell and CIE systems, as aids to expressing remotely sensed data in the most visually stimulating way possible, with the least expenditure of computing time. One method that simplifies the Munsell system to three measures of colour—hue, saturation and intensity—has received particular attention. These approximate Munsell's hue, chroma and value in a conical colour space derived mathematically from red-green-blue colour space. This transformation allows interesting means of enhancing colour and combining unrelated types of data. It is discussed further in Chapter 5.

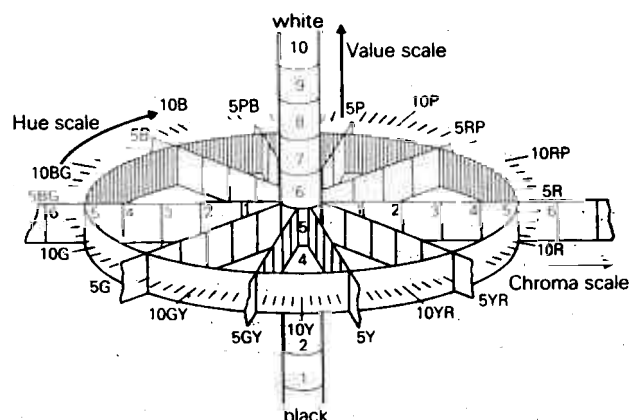


Fig. 2.15 A schematic representation of the axes used by Munsell in devising his *Book of Colour* shows the divisions that he used for hue, chroma and value. The value axis is a grey scale.

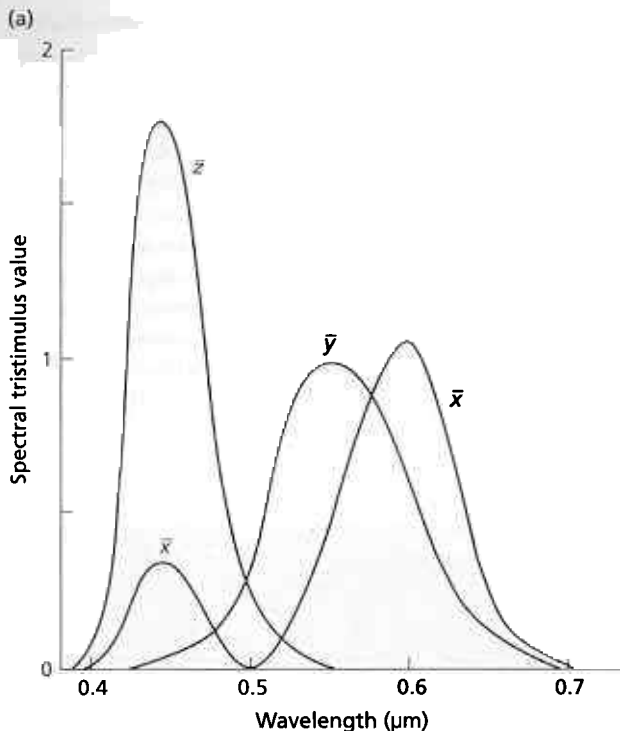
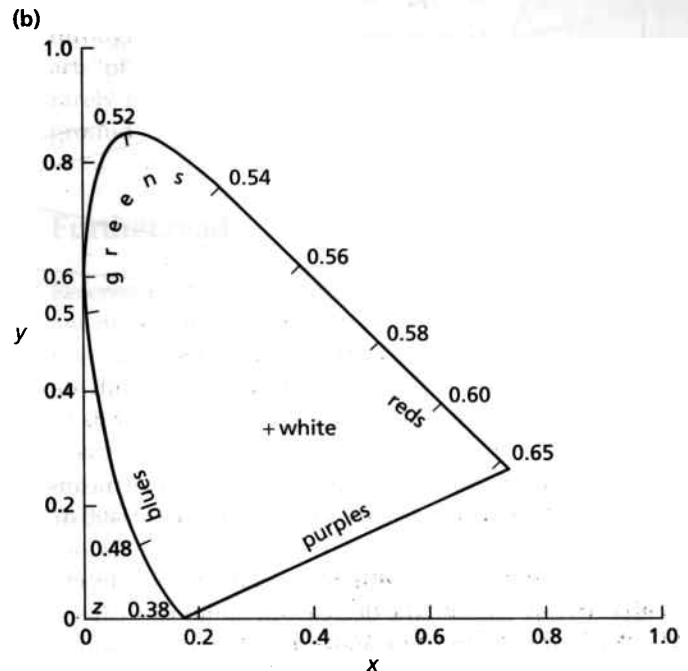


Fig. 2.16 (a) These curves represent the variation with wavelength of tristimulus values for the imaginary primaries used in constructing the Commission Internationale de



l'Éclairage (CIE) chromaticity diagram. (b) The CIE chromaticity diagram is based on plots of the tristimulus values x and y in (a) for any given luminosity.

2.5 Perception of depth

After colour vision, the most important aspect of the human visual system is based on the fact that it is binocular. A visible object subtends a perceptible angle at the retina of one eye which depends on the physical size of the object and its distance. The remembered size of a familiar object allows the brain unconsciously to estimate roughly how far away it is. For normally sighted people, an image focused on the retinas of both eyes enables them to perceive a more exact impression of the position of the object in three dimensions. The fundamental reason for this is the 6–7 cm horizontal separation of the eyes, known as the eye base.

Rays extending from a point on an object to the eyes are separated by an angle which decreases as the distance of the object increases (Fig. 2.17). The angle in radians is approximated by the ratio between the eye-base and the distance to the object. That angle is the object's absolute parallax at which the eyes must be converged to focus the object at or near the fovea. The visual cortex unconsciously fuses the two images, the outcome being a stereomodel of the location of the object in three dimensions. Because the retina can resolve only angular separations greater than about 1 minute of arc, there is a limit to stereoptic vision.

If a pencil is held at about 30 cm away, in front of a more distant object, alternate winking of each eye gives the familiar impression that the pencil moves relative to the distant object. Holding the pencil at arm's length reduces the apparent displacement. Alternately focusing on the pencil and the distant object with both eyes gives the stereoptic impression that the two objects are indeed separate distances away. With one of the objects stereoptically fused, the points of focus of each object are displaced on the retina. This is why we see two images of the distant object when we focus the pencil at the fovea. Figure 2.17 shows that the image of the more distant object falls further from the fovea in the right eye than it does in the left eye. To focus the distant object at the fovea requires a change in the eyes' angle of convergence. The difference between the two angles of convergence is the relative parallax of the two objects. It is this difference in parallax that is the most important stimulus for depth perception in the nearby environment. As in the case of estimation of distance, it can function only for angles greater than 1 minute of arc. This binocular stereopsis is the basis for three-dimensional viewing of pairs of remotely sensed images in which large relative parallaxes are produced by the method of capturing images. This is explained in Appendix A, and used extensively in the book, Chapter 4 especially, using different viewing devices.

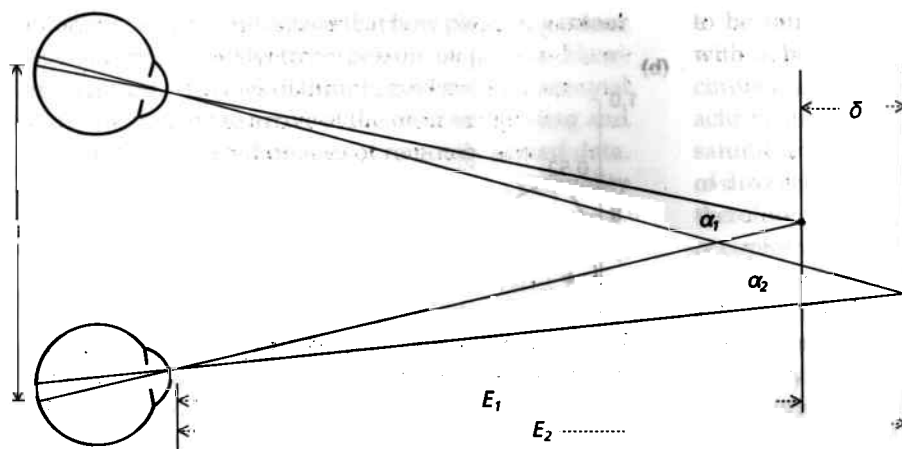


Fig. 2.17 Points on two objects at different distances have different absolute parallaxes (α_1 and α_2) measured in radians as α/E . The relative parallax is the difference between the two angles.

The limited angular resolving power of the eye means that we cannot see stereoptically beyond about 400 m. We learn to estimate distance beyond that from various features, such as the expected sizes of familiar objects, perspective effects and the decrease in luminosity caused by increasing absorption and scattering in the atmosphere. The absence of this last factor is responsible for the 'unreal' appearance of photographs taken on the Moon, where there is no decrease in luminosity of objects with distance. As most remotely sensed images represent a view from vertically above the Earth's surface, many inherent means of depth perception are of little use. One that is important is the interpretation of shadows—or misinterpretation of naturally dark surfaces—according to psychological rules that have evolved in the human visual system.

Shadows result when illumination is from the side of an object. Much of the texture in a view of the Earth's surface results from shadows cast by hills into valleys. Shadows give a very real impression of depth, but unless the direction of illumination is known the impression is ambiguous. Figure 2.18 contains perspective clues suggesting an oblique view of a landscape with circular features. Most observers will interpret the features as craters in the surface. Inverting the figure gives the opposite impression. Staring at the figure rotated through 90° for a while enables both alternatives to seem possible, although for most of the time no depth is seen. Two interesting observations can be made. In the last experiment the perspective effect, the shading and the strange orientation of the 'horizon' cause visual dissonance. The first two experiments, however, suggest that there is a subconscious assumption that illumination is from the top of the figure. Strangely, an even more convincing illusion of depth results from oblique shading. Cartographers have long exploited this by shading contoured topographic maps as if they were illuminated from the top left, or north-west on a Mercator projection. Remotely sensed images usually show no perspective,

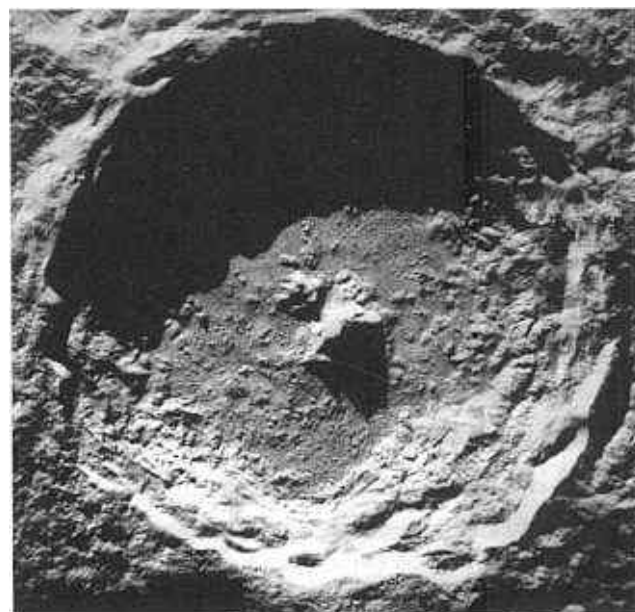


Fig. 2.18 Viewed this way up the image is clearly that of a crater with a central peak. Viewed from the top of the page the image should 'flip' from reality to inverted relief.

and shadows allow a moderately good impression of topographic relief. Many of the examples in later chapters can show both 'positive' and 'negative' topographies, however, depending on how they are viewed. In images from satellites, such as Landsat or SPOT (Chapter 3), the time of capture is mid-morning, with illumination from the south-east in the Northern Hemisphere. This invariably gives the illusion of inverted relief in correctly orientated images.

2.6 Dangerous illusions

As well as the pseudostereoscopic tricks that shadowed images can play on the observer, certain aspects of our

visual system's handling of spatial information can cause problems in image interpretation. There are hundreds of more or less well-known optical illusions and artistic tricks, which are discussed in text books on human vision. Most are based on fairly simple, sharply outlined geometric figures. Remotely sensed images are infinitely more complex, but the peculiarities of vision involved in illusion may still play a role. As described in Section 2.1, simple, complex and hypercomplex receptive fields in the visual cortex perform the useful tasks of detecting lines, orientations and angles. They help link together disparate parts of an image, thereby enabling its component objects to be interpreted. Artists have known for millennia that human vision has a predilection for triangles and slanted lines, and incorporate them in paintings to add dynamism. For the world of the familiar this is fine. However, the eye can 'invent' lines. It can also link separate points in smooth curves, and is prone to sensing the unfamiliar hidden in a familiar background—as children, we have all stared at complicated wallpaper and seen dreadful beasts lurking therein. Similar phantasms lie in wait for the unwary image interpreter.

Remotely sensed images are, by and large, less familiar than images of our immediate surroundings. They contain information relating to the unknown, or at least the partly known. The positions from which they are acquired are totally outside our normal experience, and many relate to invisible parts of the spectrum. All contain a huge variety of detail, with both organized and random elements, and therein lies the main source of illusions. Our visual system can help us to abstract meaningful spatial information, but it also can overestimate the importance of those kinds of features to which it is attuned—lines in particular. Moreover, there are suspicions that any interpreter perceives regular geometric patterns in objectively homogeneous images. There has been little research in this area. Some studies seem to indicate that the same image excites different responses in the spatial domain for different interpreters. Real features may be there, but some of those identified by the different subjects are probably spurious. Although remote sensing with image interpretation is probably

the single most efficient source of geological information, there will never be a substitute for verifying reality through field work by experienced geologists. The 'black arts' of image interpretation mean that such verification rarely needs to cover more than a small fraction of the ground.

Further reading

References to the psychophysiology of vision, so important in presenting data as images for visual interpretation, are replete within the fields of psychology. Many are difficult to understand or contain only small sections of direct relevance here. The following references are useful guides.

- Buchanan, M.D. (1979) Effective utilization of colour in multi-dimensional data presentation. *Proceedings of the Society of Photo-optical Instrumentation Engineers* 199, 9–19.
- Burns, K.L. & Brown, G.H. (1978) The human perception of geological lineaments and other discrete features in remote sensing imagery: signal strength, noise levels and quality. *Remote Sensing of Environment* 7, 163–167.
- Burns, K.L., Sheperd, J. & Berman, M. (1976) Reproducibility of geological lineaments and other discrete features interpreted from imagery: measurement by a coefficient of association. *Remote Sensing of Environment* 5, 267–301.
- Cañas, A.A.D. & Barnett, M.E. (1985) The generation and interpretation of false-colour principal component images. *International Journal of Remote Sensing* 6, 867–881.
- Cornsweet, T.N. (1970) *Visual Perception*. Academic Press, New York.
- Halliday, T. (ed.) (1995) *The Senses and Communication*. Book 3 in Biology, Brain and Behaviour (SD206). The Open University, Milton Keynes.
- Padgham, C.A. & Saunders, J.E. (1975) *The Perception of Light and Colour*. G. Bell and Sons, London.
- Siegal, B.S. (1977) Significance of operator variation and the angle of illumination in lineament analysis on synoptic images. *Modern Geology* 6, 75–85.
- Stroebel, L., Todd, H. & Zakia, R. (1980) *Visual Concepts for Photographers*. Focal Press, London.
- Wise, D.U. (1982) Linesmanship and the practise of linear geo-art. *Geological Society of America Bulletin* 93, 886–888.

3 *How Data Are Collected*

Up to the mid-1970s, geologists used remotely sensed images acquired by photographic cameras mounted on aircraft. Electronic data-gathering methods developed rapidly for two main reasons. First, the increased emphasis on planetary monitoring by unstaffed satellites and probes meant that it was impossible to retrieve film, except by costly return and re-entry manoeuvres. This demanded systems that would transmit data using microwave signals. Second, ventures into that part of the electromagnetic spectrum where film will not respond required the development of new kinds of instruments based on electronic imaging. Some of them, such as video cameras, collect and transmit images that correspond to a photographic 'snap shot'. They consist of representations of the intensity variations in a scene in the form of modulated signals similar to those used for transmitting and recording television broadcasts. The signals corresponding to radiance are continuous and produce pictures that are described as analogue images in much the same way as are photographs. They are, however, by no means perfect representations of the continuum of reality on the ground, one reason being that they are dissected in a similar fashion to images collected by the human eye. In photographs the dissection is based on the grain of film emulsion, and is not regular in detail. In electronic devices the dissection is regular by virtue of how the means of detecting radiation are structured.

Electronic systems gather data as a sequence of small parts of a scene. The data have a structure similar to that of a tiled floor or a regular mosaic, which is reconstructed to form a picture. Individual 'tiles' are known as picture elements, or pixels. Electronic, analogue devices—TV and video cameras before the digital broadcasting revolution—produce tiled images that are reproduced on a video monitor by the sweeping or 'refreshing' of the screen by electron beams. The refreshed display of the tiled matrix of pixels is termed a raster. Precise analogue measurements, such as 1.763 mV, are cumbersome to transmit, however, and the information that the signals carry is easily degraded by interference or by copying at relays. Rescaling and transmitting these numbers expressed by binary arithmetic—digital numbers (DN)—largely removes these problems. In binary a string of zeroes and ones, or 'offs' and 'ons' can express any number, so that the signal becomes as simple as possible. This means that, provided the signal is detectable, the binary digits are unmistakable. There is no degradation and the data can be copied or relayed infinitely. The range used most commonly has eight binary bits (a binary bit is a single

'on' or 'off'—1 or 0—in a binary number) or one byte, representing a range of intensities by digital numbers from 0 (00000000 in binary) to 255 (01111111). These are 8-bit integers, but any number, floating-point or complex, can be expressed in binary form too as 16-, 32-, 64- or 128-bit 'words'. Binary data, of course, are those that digital computers were designed to manipulate, and digital image processing (Chapter 5) increases analytical and interpretative power enormously.

Recent developments do not imply that aerial or satellite photography is redundant. There are vast archives of systematically gathered photographs, and they will continue to grow as long as photography maintains advantages of cost, flexibility and resolution over digital image capture. At the time of writing photography is capable of finer spatial detail over larger scenes than commercially available digital systems. The latter need to be flown at lower elevations than conventional cameras to match their resolution, with a reduction of area of coverage. Photographic film also can be scanned digitally to render it useful in image processing of various kinds. Perhaps sooner rather than later, photography will be replaced completely by digital images as technology develops. One of the pressures for that is the question of storage. Photographs are bulky and ultimately deteriorate, whereas digital images can be kept indefinitely on optical storage media that progressively take up less and less space—hundreds of frames on a single CD-ROM. The other reason for a long life for the photographic, analogue image is that approaches for the interpretation of aerial photography forms the historic basis for geological analysis of most other forms of imagery. Chapter 4 focuses on photogeology as an introduction to the interpretation of other kinds of image.

In this chapter, Sections 3.1–3.9 give outlines of the principles and techniques for each major method of data gathering used in remote sensing. Section 3.10 gives information about the different platforms from which images have been acquired and the characteristics of the images themselves.

3.1 Photography

The earliest means of remote sensing, photography, is also the most enduring and still plays an important role in geological applications. Of all methods of gathering data from a distance it is by far the most simple. All that is required is a light-proof box, a lens and shutter system, and film coated with an emulsion sensitive to radiation.

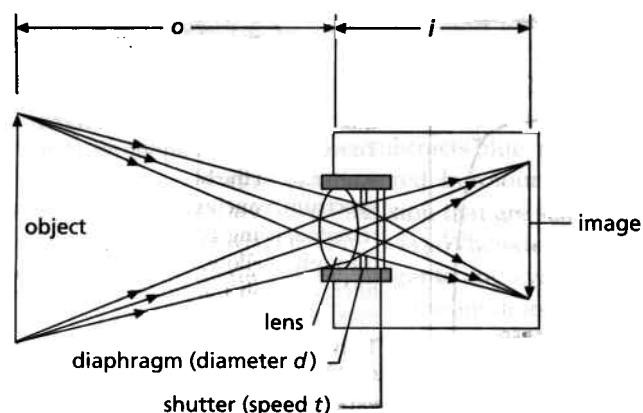


Fig. 3.1 Refraction by the glass in a lens enables rays from an object to be focused on film through a much larger aperture than would be possible with a primitive pin-hole camera. More energy is available to sensitize the film and so exposure time can be much shorter. Moreover, the aperture and exposure time can be varied using an iris diaphragm and a timed shutter. These further refine the flexibility of a lens camera, as does the provision of films with different speeds or sensitivities to light.

In a camera, radiation reflected from an object is refracted by the lens to form an inverted image on the film (Fig. 3.1). The most important attributes of the lens are its focal length and its diameter or aperture (for special applications, such as the use of stereoscopic aerial photographs in extracting elevation data—Chapter 8—the precise geometry of the lens, or the camera model, is needed). The aperture is changed using a diaphragm, and so controls the radiation-gathering capacity of the lens. How much radiation falls on the film is also governed by the speed of the shutter—the length of time the film is exposed. All cameras have shutter speeds in a conventional sequence of fractions of a second, decreasing by factors of two— $1/2$, $1/4$, $1/8$... $1/250$, $1/500$, $1/1000$. The energy E falling on the film during exposure is given by:

$$E = sd^2t/4f^2 \quad (3.1)$$

where s is the radiant energy of the scene ($\text{W mm}^{-1} \text{ s}^{-1}$), d the aperture (mm), t the exposure time (s) and f the focal length of the lens (mm). Only t and d may be varied on the camera. The aperture is expressed conventionally as the ratio between focal length and aperture diameter, known as f -stop. It is arranged in a sequence of fixed settings that complement the shutter speed, so that a range of combinations of speed and aperture giving identical exposure of the film to radiant energy can be chosen. As the energy is inversely proportional to the square of the f -stop, to achieve this a halving of exposure time must be matched by an increase in f -stop of $\sqrt{2}$ or 1.4 times. F -stops are then in the order $\sqrt{2}$, 2, $2\sqrt{2}$, 2^2 , or $f/1.4$, $f/2$, $f/2.8$, $f/4$, etc., familiar to all photographers.

In a normal camera, objects at different distances o must be focused on the film by changing the distance from lens to film i , governed by the relationship:

$$1/\text{focal length} = 1/o + 1/i \quad (3.2)$$

In remote sensing, o is of course such a large distance that focusing is unnecessary, the film being located at the focal plane of a fixed lens system.

In all remote-sensing systems, two attributes are paramount: the minimum energy and the smallest object that can be detected on an image. The two are interrelated. The minimum separation between two objects at which they can be distinguished—usually after enlargement—is the spatial resolution. It is often described by the largest number of lines per millimetre that can be seen in the image, which can be related to dimensions on the ground using the optical geometry of the system. The faster the film, the coarser the resolution, and commonly used film ranges from ISO 8 to 640, with resolutions from 250 to 25 lines mm^{-1} , which correspond to ground resolutions of 6–60 cm at an image scale of 1 : 15 000. In practice spatial resolution is rated using test charts with high-contrast dark and light lines at different spacings. A better measure is the system's response to a point source of energy, such as a star. This will be blurred and spread out to a greater or lesser extent, forming a pattern known as the point-spread function (Section 2.2, Fig. 2.8).

In photographs limits on resolution are posed by the quality and size of the lens, and are affected by defects such as spherical and chromatic aberration. Both affect the sharpness of focusing. Chromatic aberration results from the variation of refractive index and therefore focal length of lenses according to the wavelength of radiation. It produces different focusing of the component radiation on the film, so giving blurring. The film itself is coated with grains of photosensitive chemicals in an emulsion, the size of the grains varying with the speed of the film. A fast film has coarser grains and lower resolving power than fine-grained, slow film, but is able to record lower energies at a particular shutter speed and f -stop. Because fast film is more sensitive, there must be a trade-off between resolving power and detecting power.

Black and white film depends on the photosensitivity of silver salts. When a photon strikes a grain in the film an electron is freed from the silver salt molecule. It migrates until it lodges at a defect in the lattice, where it converts a silver ion into a metallic silver atom. This triggers a chain reaction converting all silver ions in the grain into metal. The larger the grain, the greater the chance of an encounter with a photon; that is why fast films have coarser grain than slow films. Developing dissolves away the unconverted silver salt leaving a negative image of black silver grains, the density of which represents the energy distribution falling on the film. Printing is the reverse of this process.

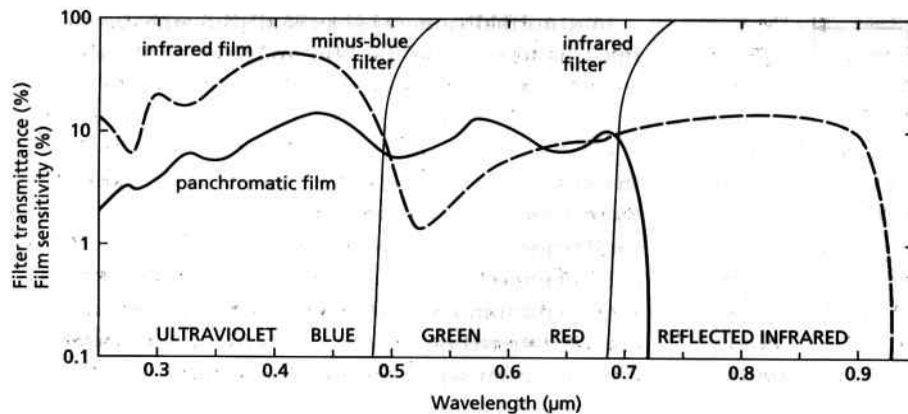


Fig. 3.2 There is considerable overlap between the sensitivities of black and white panchromatic films and those sensitive to reflected infrared. To remove the swamping effect of scattering in the blue part of the spectrum, panchromatic film used in remote sensing generally is exposed through a yellow, or 'minus-blue' filter. This only transmits wavelengths to the right of its transmittance curve. To remove the effects of reflected visible radiation, infrared film is exposed through a dark red filter.

There are two main types of black and white film used in remote sensing, those sensitive to radiation with wavelengths up to $0.7 \mu\text{m}$, covering the whole visible part of the spectrum—panchromatic—and those with a sensitivity that extends to about $0.9 \mu\text{m}$, including part of the near-infrared region (Fig. 3.2). Infrared-sensitive film was devised originally as a means of distinguishing between vegetation, which strongly reflects VNIR radiation (Section 1.3.3), and green camouflage netting or paint, which does not.

As explained in Section 1.3.1, the blue end of the spectrum is dominated by atmospheric scattering and haze. To improve the contrast of panchromatic photographs they are normally exposed through a yellow, blue-absorbing filter, and are known as minus-blue photographs. Similarly, to reduce the effects of visible light, infrared film is usually exposed through a dark-red filter (Fig. 3.2), which absorbs blue and green light. Examples of both types of image are shown in Fig. 3.3.

Whereas the human eye and colour television operate

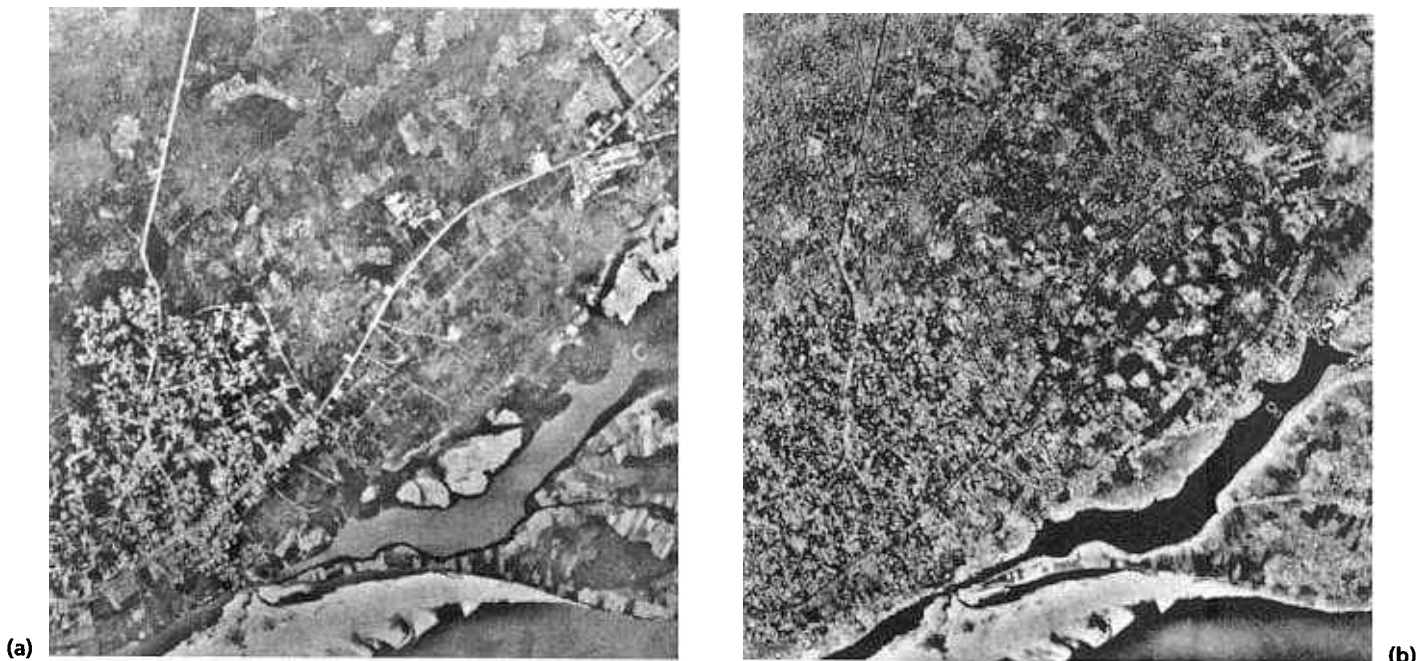


Fig. 3.3 This pair of 1 : 40 000 minus-blue panchromatic (a) and infrared (b) aerial photographs are of the same area on the coast of West Africa. They were taken under heavy cloud cover and are dimly lit as a result. The panchromatic image shows most of the details of the township, roads and the coastal sand spit, but contains very little information relating to vegetation. The infrared image has higher contrast, and is an excellent means of discriminating vegetation, although the cultural features are poorly displayed. The silt-laden waters of the

lagoon show in black, because water strongly absorbs. This makes the water courses more clearly visible on the infrared image. Vegetation is a strong reflector of near-infrared radiation, and most of the variation in the fields is the result of crops being at different stages of growth. Individual trees, especially the mangroves near the lagoon, show up as bright speckles. The panchromatic image in vegetated areas shows bare soil as light tones and crops and trees as a monotonous dark grey. Copyright Aerofilms Ltd.

with additive primary colours—red, green and blue—colour film uses subtractive colour, for which the primaries are magenta, yellow and cyan (Section 2.4). Each results from subtracting one of the additive primaries from white light. A yellow filter subtracts blue, magenta subtracts green and cyan subtracts red. In colour positive film there are three layers of emulsion that are sensitive to blue, green and red light (Fig. 3.4a). When developed these emulsions retain yellow, magenta and cyan dyes, respectively, in inverse proportion to the amounts of blue, green and red light to which they have been exposed. White light passing through the developed film has its primary additive colours filtered by the dye layers to produce a positive colour image. As an example, developing a film that has been exposed to blue light bleaches the yellow layer completely but does not affect the magenta and cyan layers. These layers filter out green and red from white light transmitted through the film, leaving only the blue component.

Colour infrared film also includes yellow, magenta and cyan dye layers, but each has a different sensitivity compared with normal colour film (Fig. 3.4b). In this case all the layers are sensitive to blue light, as well as to green, red and near-infrared light, respectively. Removing the swamping effect of blue means placing a blue-absorbing, yellow filter in front of the lens.

The appearance of natural colour photographs needs no amplification, except to mention that their fidelity may vary with different proprietary materials and developing processes. Colour-infrared photographs, however, express infrared, red and green radiation as red, green and blue, respectively, and so have an unusual appearance (Plate 3.1). They are examples of what are termed false-colour images. In their case, vegetation which reflects VNIR strongly and absorbs a high proportion of visible light appears as various shades of red and magenta. Red materials, such as haematite-stained sand, reflect infrared strongly too. The combination of weak blue, intense

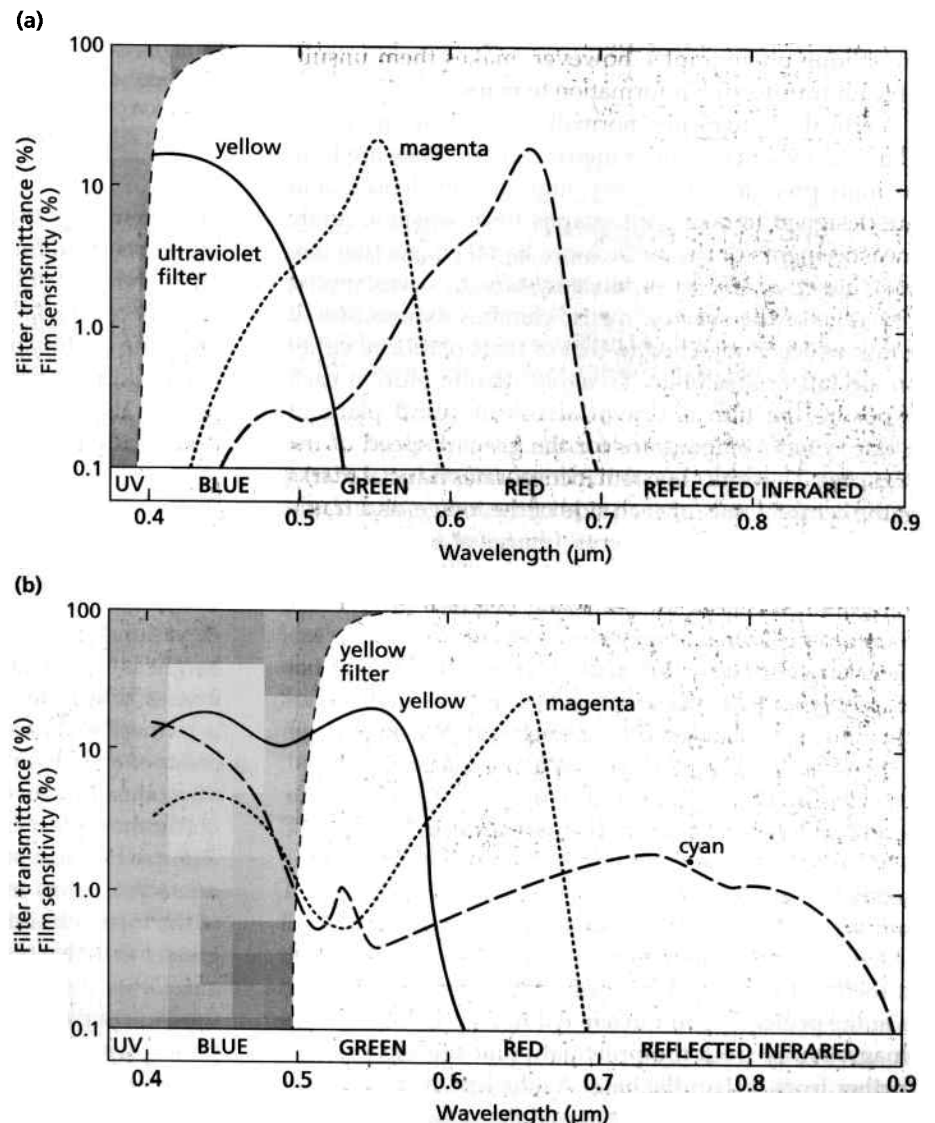


Fig. 3.4 The ranges of sensitivity of the yellow, magenta and cyan dye layers in natural colour (a) and colour-infrared (b) positive films are different, as are the ranges of transmittance of the filters used with them.

green and intense red in the false colour image imparts a yellow or orange tint to such surfaces (Section 2.4). Since water absorbs infrared totally, damp soil frequently appears in blue or cyan hues (Plate 3.1).

The interpretation of panchromatic, infrared, natural colour and false colour photographs forms an important topic in later chapters. However, they are limited to that part of the spectrum with wavelengths shorter than $0.9\ \mu\text{m}$. Investigation of longer wavelengths requires electronic image gathering devices. Before discussing them, some of the geometrical attributes of photographs need to be introduced.

Photographs of the Earth's surface may be taken from any angle, but the most useful are those taken looking vertically downwards to give a map-like view. Oblique photographs are sometimes taken, and they have two advantages: they show perspective, which allows the brain to estimate distance and scale; and they display outcrops of low-angled planar structures that are difficult to pick out in vertical images because they follow irregular topography closely. The rapid changes in scale in oblique photographs, however, makes them unsuitable for transferring information to maps.

Vertical photographs normally are taken one frame at a time by a mapping or metric camera. They are built to high precision with very high quality lenses, and are designed to take photographs from which accurate measurements of the surface can be taken. As they are used most widely to produce systematic, overlapping coverage of the surface, metric cameras expose film at a rate which is matched to that of their platform, either an aircraft or a satellite. To avoid motion blur, at each exposure the film is drawn across the focal plane at a rate which compensates for the ground speed of the platform. The optical system incorporates fiducial marks at the centre points of each side of the image, as a frame of reference for measurements. Images of an altimeter, a spirit level, a clock and a counter on the photograph's margins help check on the flying height and attitude, time and position in sequence.

A vertical photograph gives a view of the Earth that we rarely experience, so at first sight it has an unusual appearance. Over a familiar subject, such as the centre of a city (Fig. 3.5), one cause of this unfamiliarity is evident at once. Tall buildings appear to lean outwards from the centre or principal point of the photograph. This radial-relief displacement is easily explained by the optical geometry involved (Fig. 3.6). Tracing rays from the tops and bases of tall buildings through the lens to the focal plane shows that the angle subtended by the building at the focal plane changes from zero when the building is at the principal point to a maximum at the edge of the image. Away from the principal point the top projects further from it than the base. As the lens is circular the same effect occurs everywhere, and the amount of relative

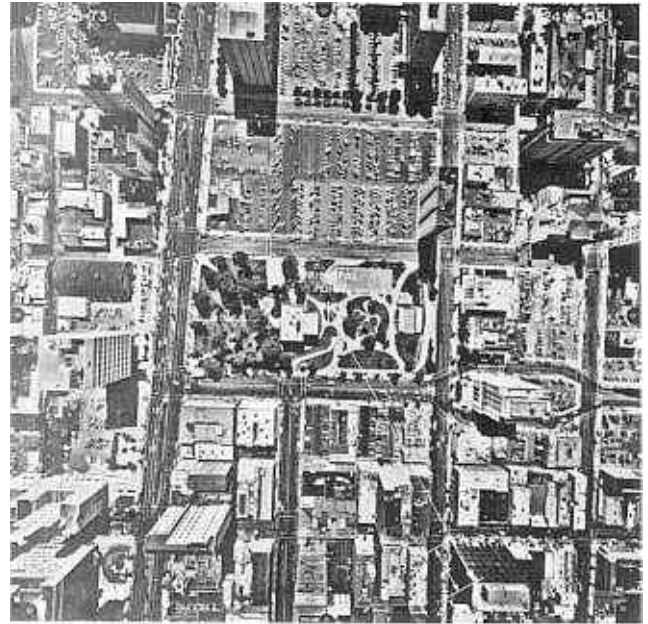


Fig. 3.5 Vertical aerial photographs of urban areas dominated by high-rise buildings show clearly the effects of radial relief displacement, and of the increase in scale with increased elevation of the surface.

displacement of the top increases radially outwards from the principal point. The wider the lens' field of view and the lower the camera height, the more pronounced the effect becomes. From the geometry of Fig. 3.6(a) and the principle of similar triangles you can see that the height of a building h relates to its distance from the principal point r , the radial displacement of its top relative to its base d and the flying height H by:

$$h = Hd/r \quad (3.3)$$

So, it is possible to make accurate measurements of elevation from vertical photographs, provided that h/H is significantly greater than zero. That holds for most elevations of any importance, provided that the flying height is not of the order of tens or hundreds of kilometres, when the displacement is very small. If the need is to measure changes of elevation of the ground of the order of a metre or less, as in some engineering applications, then low-level flights are necessary.

Another effect of the variations in height of surface features that can be seen on Figs 3.5 and 3.6 is that they cause variations in scale on the image. The separation of the tops of buildings A and B on Fig. 3.6 is noticeably greater than that of their bases, whereas both are equidistant. Other distortions result from different attitudes of the camera platform (Fig. 3.7) that cause slightly oblique photographs, the scale of which changes across the image. Scale also may change from one photograph to another in a sequence because of variations in flying

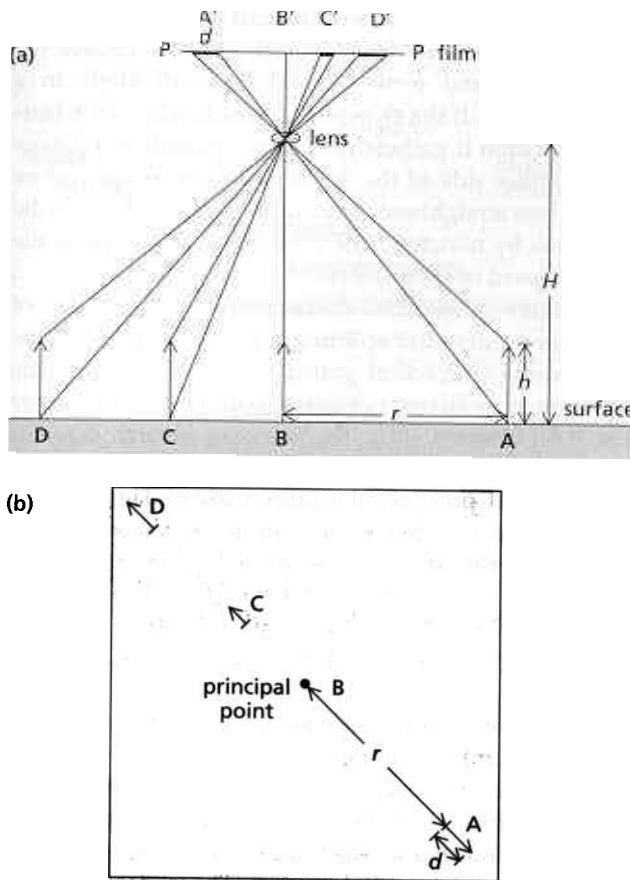


Fig. 3.6 Tracing the ray paths from the tops and bases of equally tall buildings through the lens of a camera to the film (a) shows how radial relief displacement is produced in a vertical photograph. The further the building is from the point directly below the lens—the nadir, which is imaged as the principal point of the photograph—the larger the angle it subtends at the film. On the film (b) the top of a building is imaged further from the principal point than its base, and both base and top are joined by a radial line to the principal point. The building therefore appears to lean away from the centre of the image (Fig. 3.5). Moreover, the scale at the level of the bases of the buildings is smaller than the level at their tops.

height. Distortions of these kinds are unavoidable and can be corrected only by computer (Appendix B).

Another defect of photographs that is systematically related to their geometry is a radial decrease in overall brightness called vignetting. This is a result of the radially increasing angle between camera lens and surface, and a corresponding decrease in the apparent aperture or light-gathering capacity of the lens. Filters in which density increases towards their centre can compensate this.

Taking vertical photographs usually involves sequences of frames along parallel flight lines. The interval between exposures is arranged to allow for sufficient overlap between images so that every point on the ground appears

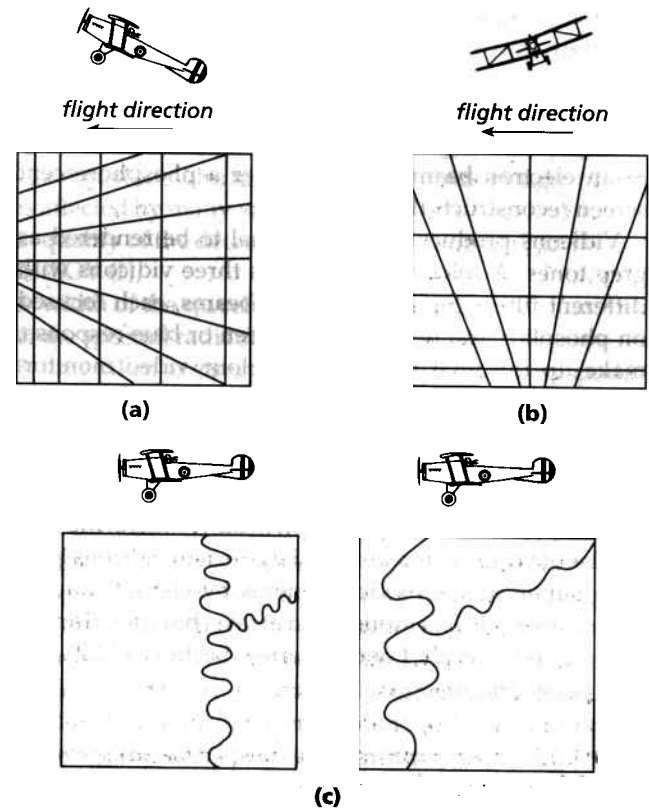


Fig. 3.7 Apart from the distortions in scale owing to varying elevation of the surface, the main distortions in aerial photographs are a result of tilting of the aircraft (a and b) so that the view is slightly oblique. A perspective effect of changing scale is introduced. Convection cells in the atmosphere may cause the aircraft to change altitude so that adjacent images have different scales (c).

on at least two separate photographs. This imparts parallax to objects with different heights. Adjacent overlapping photographs represent the views by left and right eyes, but with the eye separation increased to hundreds of metres or even several kilometres. Viewing overlapping photographs with appropriate binocular lens or mirror systems simulates stereoscopic vision with an exaggerated three-dimensional view of the ground. Such stereopairs form the basis for both manual and digital elevation measurements (Appendix A).

3.2 Vidicon cameras

Television originally used a camera known as a vidicon. Like a photographic camera, a vidicon gathers radiation through a lens system and passes it through various filters to focus on a flat target. Instead of a photosensitive emulsion, the coating on a vidicon target is a transparent photoconductive material. The electrical conductivity of the target increases with the intensity of the illuminating radiation. Scanning the target with an electron beam as

a series of lines builds up the picture. This maps the variation in conductivity, and so the illumination, as an analogue electrical signal. The time involved is around that involved in photography, so the signal is effectively a snapshot. Using the signal to modulate the intensity of an electron beam sweeping over a phosphorescent screen reconstructs the image.

Vidicons produce only one signal to be rendered as grey tones. A colour system needs three vidicons with different filters. Separate electron beams, each focused on phosphor grains giving red, green or blue response, make up the final image on a colour video monitor. Like film emulsions, the materials used in vidicons are sensitive only to visible and near-infrared radiation. This, together with the physical limitations on resolution imparted by the size of the vidicon target and by the accuracy of the scanning electron beam, restricts the usefulness of vidicon cameras. To a large extent vidicons have been superseded in video cameras by charge-coupled devices (Section 3.4) mounted in arrays, but the principle applied to the early Landsat series (Section 3.12.2) and interplanetary missions.

3.3 Line-scanning systems

A similar effect to the raster from a vidicon comes from using a single photosensitive detector cell to scan radiation emanating from a series of strips of the Earth's surface. One means of doing this involves a line-scanner. The forward motion of the platform allows successive

strips of the terrain to be monitored. A mirror directs radiation on to the detector and a motor rotates the mirror back and forth through a small angle in a repeated cycle. If the sweep is perpendicular to the platform's motion it gathers data from immediately below and to either side of the flight path. This builds up an image from straight swaths or lines orientated across the flight line by matching the frequency of sweeps to the ground speed of the platform.

Direction of radiation by the mirror on to banks of detectors occurs after splitting it into a series of wavebands using diffraction gratings. This means that the detectors sense all the chosen wavebands simultaneously (Fig. 3.8). Consequently, the response of each detector corresponds to exactly the same portion of terrain as all the others, at precisely the same instant. This ensures that the data for each waveband are in exact register. To improve resolution one sweep of the mirror directs radiation to several detectors for each waveband. The swath of finite width on the ground comprises several strings of signals, each of which represents a smaller line within it.

Line-scanners escape the spectral limitations (between 0.4 and 1.1 μm) of photographic and vidicon methods. Radiation-sensitive detectors now span the spectrum from ultraviolet through to the region where energy emitted by the Earth reaches a peak (Fig. 3.9), although no single detector covers the whole range. Those detectors for radiation beyond 1.1 μm give measurable responses only when cooled, either by liquid nitrogen or by exposure to the very low temperatures of outer space.

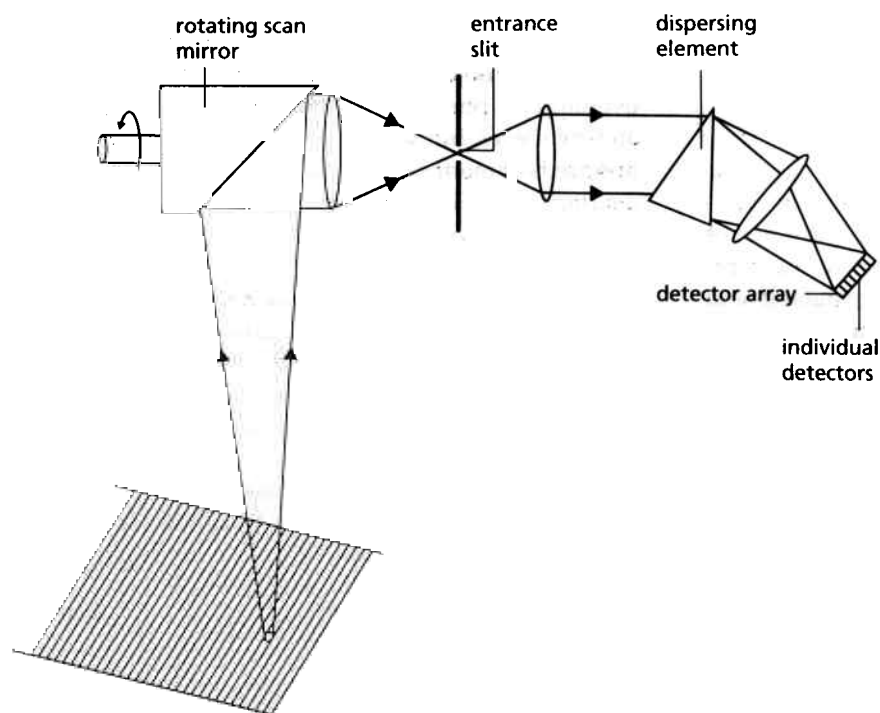


Fig. 3.8 Radiation gathered by a scan mirror in a line-scanner is focused by an optical system through a dispersion system, either a prism or a diffraction grating, on to banks of detectors. There is at least one detector for each spectral band being sensed, and in most cases there are several so that one sweep of the mirror across the surface produces a number of lines of data for each band.

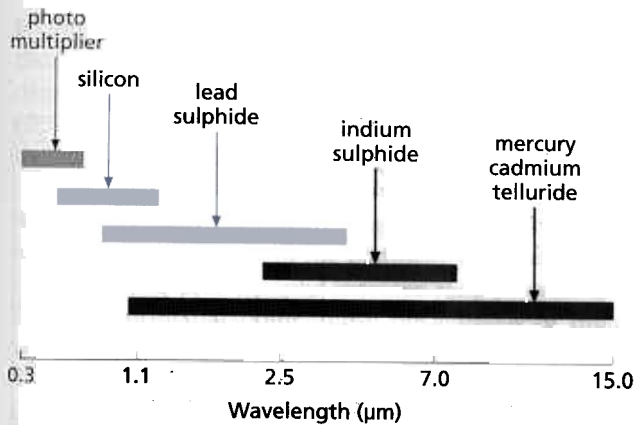


Fig. 3.9 Several photoelectronic detectors cover the whole of the useful range of visible and infrared wavelengths. Each has an optimum response over a relatively broad range. To produce data from narrow wavebands they must be used in conjunction with filters or be placed precisely in a spectrum produced by a diffraction grating. Except for photomultipliers or silicon diodes, most detectors need to be cooled.

Line-scanner resolution is analogous to that of a photographic camera, but is a function only of the geometry of the optical system and the sensitivity of the detectors. At any instant, the area on the ground from which radiation is focused on a detector depends on the focal length of the optics and the flying height. This measure of resolution is the instantaneous field of view (IFOV), defined as a solid angle at the apex of a cone, within which all radiation is intercepted by the detection system. Expressed as dimensions on the ground, the IFOV is:

$$d = h/I \quad (3.4)$$

where d is the diameter of the area on the ground, h is the flying height and I is the IFOV expressed as a cone angle. This ground area, the time taken to focus energy from it on the detectors and the size of the scanning mirror determine the total energy that is available for detection at all wavelengths. The system's ability to discriminate different energy levels is controlled by the sensitivity of the detectors. Sensitivity forms the main practical limit on resolution. Whereas it may be technically feasible to have a very small IFOV, the detectors would show only broad variations in radiance from the surface. The image would be of poor quality. The broader the range of wavelengths directed on to a detector the more energy is available for it to detect. So for a small IFOV, the broader the waveband, the better is the discrimination of different signals from the surface. With a large IFOV, a narrower waveband can be monitored and still discriminate different signals from the ground.

Any detector experiences random fluctuations caused by electronic or structural defects in its design or power supply. To a greater or lesser degree all produce a response to this noise in the absence of any incident

energy. This generally appears as speckles on an image of an otherwise uniform surface. In practice the performance of a detector is assessed by its signal-to-noise ratio (S/N). Noise is generally constant and so there is a minimum signal, below which distinction is not possible. The signal-to-noise ratio at a particular wavelength (S/N) _{λ} is affected by many variables:

- 1 quality of the detector at that wavelength (Q_λ);
- 2 IFOV (I);
- 3 width of the waveband being detected ($\Delta\lambda$);
- 4 the spectral radiant flux of the surface (E_λ);
- 5 time taken to sweep the portion of surface corresponding to a pixel (t).

The variables express (S/N) _{λ} empirically by:

$$(S/N)_\lambda = Q_\lambda \times I^2 \times \Delta\lambda \times E_\lambda \times t \quad (3.5)$$

This empirical relationship underlies the design of line-scanners.

Several options help improve, i.e. increase the signal-to-noise ratio, without having to develop a better quality detector. These are: increasing the IFOV (coarsening the spatial resolution), widening the waveband, or increasing the time over which energy is collected for each pixel (slowing the scan speed of the mirror, or the velocity of the platform). Narrowing the waveband to detect small spectral features is possible only by increasing the IFOV, or by increasing the time taken to sweep each pixel. This is important because many of the spectral features of interest shown in Chapter 1 are extremely narrow, and a small $\Delta\lambda$ is needed to resolve them.

Clearly there are trade-offs in trying to improve spatial and spectral resolution while keeping an acceptable S/N . The most important limiting factor is the time dwelt on each pixel, because this involves the velocity of the platform and the need to cover all parts of the surface. Spacecraft have a fixed orbital velocity, depending on altitude, and there is a lower limit for aircraft too—their stalling speed. Moreover, the performance of the detectors is not constant with wavelength. The spatial and spectral resolutions, and quality of data from operational line-scanners have to be optimized.

There are two methods of recording line-scan data, in analogue or digital form, either on magnetic tape aboard the platform or after telemetry down to a ground receiving station. In the analogue case the signal modulates a light source that moves in lines across a photographic film, so building up an analogue image in the form of a strip. More usually the recording is in digital format. The first step involves dividing the continuous signal from the detector into discrete blocks, either mechanically using a rotating shutter blade, or electronically using a timer. In either case the block of signals correspond to finite lengths of recording time and thus to distances along the scanned swath. The signals in each block are averaged, rescaled and converted to a digital number

(DN), which is recorded. Each DN represents the radiant energy in a particular waveband associated with a pixel on the ground. An image recorded in digital format therefore comprises a geometrically precise array of pixels in lines that is identical in structure for each waveband, so that they can be registered exactly. This array, or raster lends itself to both computer manipulation and to video display.

The size of each pixel is a function of IFOV and flying height. Although the resolution of the system is defined as the finest spacing of objects on the ground that can be discriminated in an image, it is the pixel size that is quoted most commonly as 'resolution'. Whether an object is detectable or not depends on its size, its contrast with its surroundings and its shape and orientation with respect to the raster pattern. Figure 3.10 shows that pixels often overlap several objects or classes of ground cover. The response for each of these mixed pixels has a contribution from each enclosed object or part of an object, depending on their size and contrast relative

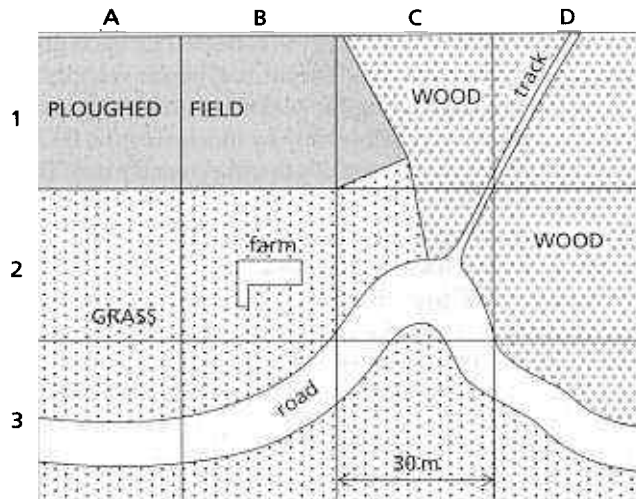


Fig. 3.10 A series of 30 m pixels superimposed on the ground surface shows that most pixels include several objects and surface categories. Only pixels A1, B1 and D2 are 'pure'. The rest are mixed pixels. If the farm had a sufficiently high contrast with grass its effects on pixel B2 could be high enough that the pixel had a different brightness from those surrounding it. The farm might be detectable, but it could not be recognized as a farm. Similarly the road could affect pixels A3, B3, C2, C3 and D3, and probably others beyond the map. Because road-containing pixels would be linked, however, the crude shape and orientation of the road could be seen. The track may be too narrow to have a noticeable effect. Pure luck has resulted in the boundary between the ploughed field and grass falling along a boundary separating four pixels. On the image it would be clear and represented accurately. All the other boundaries fall within pixels. Although they would contribute to the brightness of these pixels, on the image the boundaries would become steps in the brightness between pixels dominated by large classes. Their position and orientation would become 'blurred' by the rectangular raster.

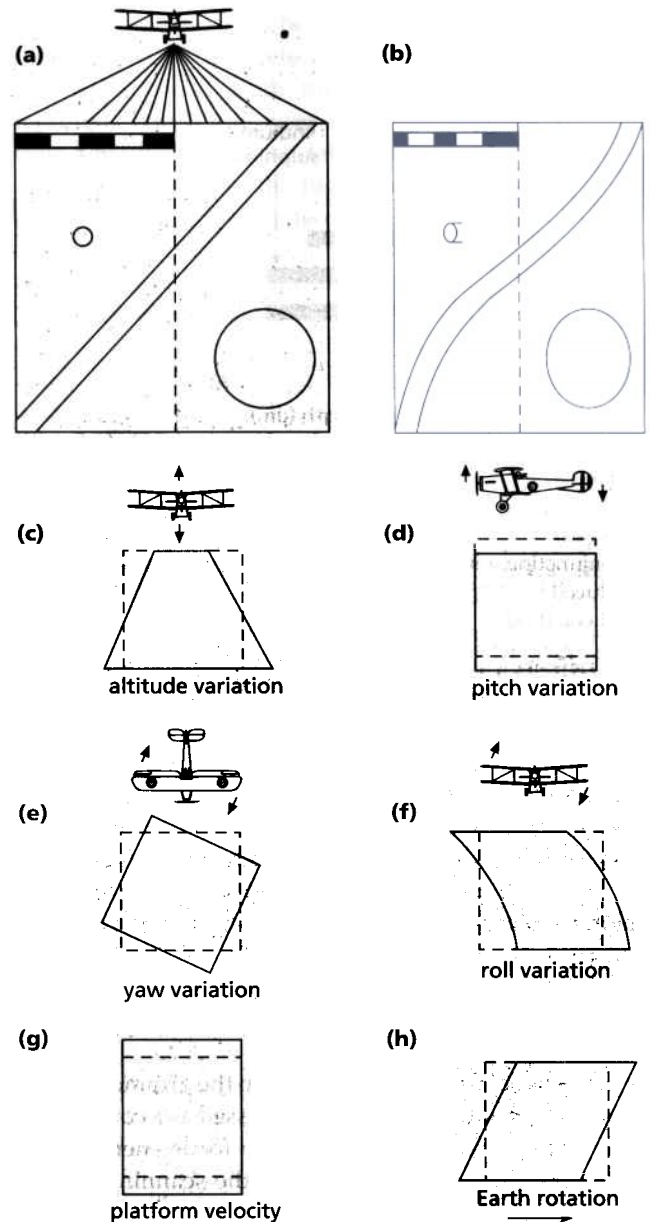


Fig. 3.11 Because the distance on the ground covered by the same angle subtended at the scanning mirror increases away from the platform (a), all line-scan images decrease in scale away from the ground track of the platform (b). The scale along the ground track does not change, so straight linear features running diagonally across the surface are distorted in a characteristic sinusoidal fashion on the image (b). Tall objects appear to lean away from the ground track and the shapes of objects become distorted. As well as this effect, distortions resulting from instability of the platform result in the rectangular format of an image (dashed outlines) representing more or less irregular or displaced areas of the surface (c to g). In the case of satellites, the Earth rotates beneath the platform as it passes over. As a result the area imaged as a series of lines scanned during a period of time represents a parallelogram on the surface (h).

to their surroundings. An object smaller than a pixel therefore may be detectable if it has a high contrast. This does not mean that it can be recognized, however, for recognition depends on shape, and shape is masked by that of the pixel itself.

Like vertical photographs, line-scan images suffer from a variety of distortions, resulting from both the optical system and the motion of the platform. Whereas optical distortions on photographs increase radially outwards from the principal point, those on a line-scan image increase outwards from the groundtrack along each line. This is because the mirror changes the viewing angle of the ground continuously during each sweep. Both scale and geometry change away from the groundtrack, so that tall objects lean in parallel away from the mid-line of the image, and scale decreases in the same fashion. Because of varying lateral scale, the apparent direction of ground features changes across the image. Straight

lines not parallel to scan lines or the groundtrack take on a curved shape as a result (Fig. 3.11). As the separation between sensor and ground increases along the scan line, so the resolution becomes coarser towards the image edge. The higher the platform and the smaller the sweep angle, the smaller these distortions become, and with some satellite systems they are barely noticeable.

Distortions resulting from platform motion are present in all line-scan images to a greater degree than in near-instantaneous photographs, because they are built up over a relatively long period of time. Platform altitude and flight direction may vary. There can also be roll, pitch, yaw and heave. The Earth rotates beneath satellites while the image is being scanned, so that a rectangular image is really a parallelogram on the ground (Fig. 3.11h). Analogue images (Fig. 3.12) can rarely be corrected for such distortions. Digital images, however, permit computerized registration to maps (Appendix B).



Fig. 3.12 This line-scan image of thermally emitted infrared wavelengths over a power station in Nottinghamshire, UK, shows distortions of the type featured in Fig. 3.11(b). They are

particularly noticeable for the road and the array of cooling towers. The light specks at lower left are cattle. Copyright Aerofilms Ltd.

The final point about defects in line-scan images stems from their structure. Any fluctuation in the response of an individual detector or miscalibration between detectors in an array for one waveband results in striping parallel to the scan lines. Sometimes, a detector may fail temporarily, or its response may go out of phase with the others in an array. These failures produce lines with no data, and pixels displaced relative to those either side, respectively. Random electrical discharges in the system may also introduce spurious numbers into an image, which show as bright pixels that relate to nothing on the ground. Appendix B covers various defects and their removal in more detail.

3.4 Pushbroom systems

Line-scanners contain mechanical components that sometimes fail and have a poor signal-to-noise ratio. Most important, they cannot have both high spatial and high spectral resolutions, because they dwell on each pixel during a sweep of the mirror for a very short time. Tiny (< 10 mm across) radiation-sensitive cells, called charge-coupled devices (CCDs) help solve the last problem. Charge-coupled devices use the same materials as those shown in Fig. 3.9. Because they respond to wavelengths up to about $2.4\ \mu\text{m}$, CCDs can monitor the full visible and near-infrared (NIR) range in which many spectral features of interest occur. A CCD becomes charged in proportion to the incident radiation, and discharges very quickly to produce either an analogue or digital

signal. In effect, a CGD is a light-sensitive capacitor. A rectangular array of CCDs forms the light-gathering device in both video and digital cameras. Arranged in thousands along a linear array on to which an optical system directs radiation from the ground, no moving parts are required to build up an image from a moving platform. The array simply and directly measures radiation from pixels along a ground swath swept by it. Such a device is known as a pushbroom system—it works in an analogous fashion to the pushing of a wide broom across the floor. Pushbrooms build up images of successive lines perpendicular to the flight path, each column of pixels corresponding to the response of each CCD in the array (Fig. 3.13a). In addition to their advantages of simplicity together with direct measurement from geometrically exact pixels, pushbroom systems are reliable and require little power. The SPOT pointable pushbroom (Section 3.12.3) can image areas to either side of its ground track by using a variably orientated mirror (Fig. 3.13b). The spatial resolution of a pushbroom system is a function of the size of each CCD, the optics and the flying height. Its spectral resolution, like that of a line-scanning system, depends on the time during which each CCD gathers radiant energy.

A satellite carries two imaging systems, one a line-scanner, the other a pushbroom. Each produces images with pixels that are 30 m square arranged in lines 3000 pixels long. One sweep of the line-scanner's mirror takes a time t , so the time dwelt on each pixel is $t/3000$. The pushbroom captures *all* 3000 pixels in the same line at once, in time t . So a pushbroom's CCDs gather

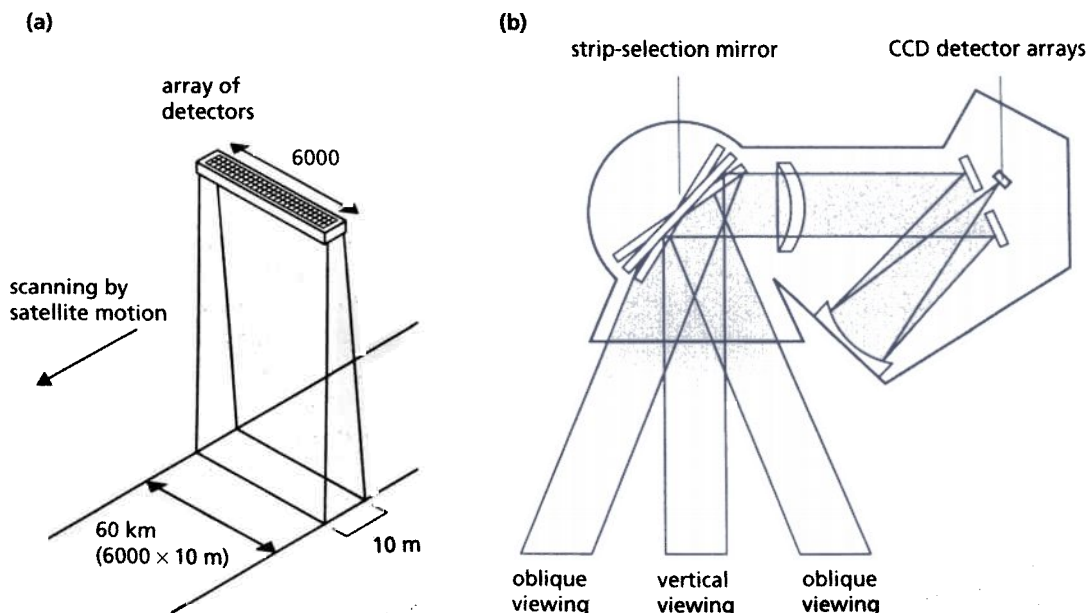


Fig. 3.13 A pushbroom system produces an image by using the motion of the platform to sweep an array of charge-coupled devices (CCDs) across the surface (a). Radiation is directed on to the array by a mirror (b), the angle of which can

be changed to allow imaging ahead of the platform or to one side. This enables stereoscopic images to be acquired and, in the case of sideways direction allows the same area to be imaged on several separate occasions without direct overlap.

radiation for 3000 times longer than do the detectors in a line-scanner. This gives three advantages for pushbroom systems. First, their signal-to-noise ratio is much better, and so they can monitor lower radiation levels. Second, they can detect smaller pixels, by increasing the number of CCDs in the pushbroom and by reducing the time of data collection for each line, thereby making it narrower on the ground. Thirdly, it becomes possible to reduce the term $\Delta\lambda$ in Equation (3.5). In other words, the spectral resolution can be improved to match the narrowness of important spectral features. The design of imaging spectrometers (Section 3.6) exploits this last option.

3.5 Microwave imaging

The longest wavelengths used in remote sensing are more than a thousand times longer (1–25 cm) than those accessible to line-scan systems ($< 15 \mu\text{m}$). This is the microwave region of the radiation spectrum. All materials above absolute zero emit microwaves, but for the ambient temperature of the Earth's surface (300 K) the energy curve (Fig. 1.3) shows that energies in this region are low. Gases and clouds hardly attenuate microwaves, however, and they can pass through several metres of dry solids. Detection of natural microwaves from the Earth, like most other techniques, is a passive method. Microwaves also can be generated electronically using klystron tubes, when coherent radiation of a single wavelength is produced. Detection and analysis of this artificial radiation when scattered back to a detector by the surface is therefore an active method, originally termed radio detection and ranging, or radar.

Passive microwave remote sensing collects data using an antenna, similar to a radio-telescope. Most surveys produce profiles of microwave emittance along lines, rather than images. Images, in the familiar form of line-scans, become possible if the antenna is swept from side to side beneath the platform. These scanned images suffer from the same distortions as those of line-scan systems operating in the visible and near-infrared regions, but their spatial resolution is very much coarser, and each wavelength requires a separate antenna.

The IFOV of a passive microwave system is expressed as a solid angle (θ), and is given by:

$$\theta = 1.2\lambda/D \quad (3.6)$$

where λ is wavelength and D is the diameter of the antenna. For microwave wavelengths and fine resolution D is too large to be portable, and except for very low flying heights the spatial resolution is coarser than with other imaging systems. For this reason, there have been few experiments in geological applications that use passive microwave technology. Its main use has been directed at climatology and oceanography.

Table 3.1 Coding of divisions in the microwave region used by radar systems

| Band code | Wavelength (cm) |
|-----------|-----------------|
| K_a | 0.8–1.1 |
| K | 1.1–1.7 |
| K_u | 1.7–2.4 |
| X | 2.4–3.8 |
| C | 3.8–7.5 |
| S | 7.5–15 |
| L | 15–30 |
| P | 30–100 |

During World War II, when radar was being developed, security meant giving letter codes to different microwave wavebands (Table 3.1). Those used most commonly are K_a , X, C and L bands.

The operation of active microwave or radar imaging systems differs fundamentally from that of all other systems. Microwave radiation 'illuminates' the surface with a downwards look direction that is to the side of the platform. It is sideways-looking radar. Producing images does not involve using an antenna to sweep a ground area; that would limit the size of the antenna and thus the resolution. Instead, a fixed antenna detects returning radiation from a long, narrow strip perpendicular to the flight direction of the platform (Fig. 3.14a).

To avoid interference between the emitted radiation and that returning from the surface, illumination involves pulses. Timing of the pulses is crucial, so that all radiation returning from the surface arrives before another transmitted pulse. The timing depends on the maximum distance to the side, or the range, from which returns are expected. The frequency of the pulses is inversely proportional to the range, because the further the range, the longer returns take to reach the antenna. Because radar travels at the speed of light, the timing is in microseconds.

The response of the antenna to energy returning from the surface from each pulse comprises both a measure of the returning energy and the time elapsed since the pulse. This time (t) is a measure of distance to the surface from which energy was returned. The distance is twice that from the platform to the object (the slant range). As radar waves travel at the speed of light (c), the slant range is $ct/2$. As Fig. 3.14(a) shows, the distance on the ground or ground range (r) is slant range multiplied by the cosine of the depression angle (α) for the radar beam at that point:

$$r = (ct \cos \alpha)/2 \quad (3.7)$$

The depression angle decreases as range increases.

A line on a radar image is therefore a plot of energy with time, resulting from the outward propagation of a

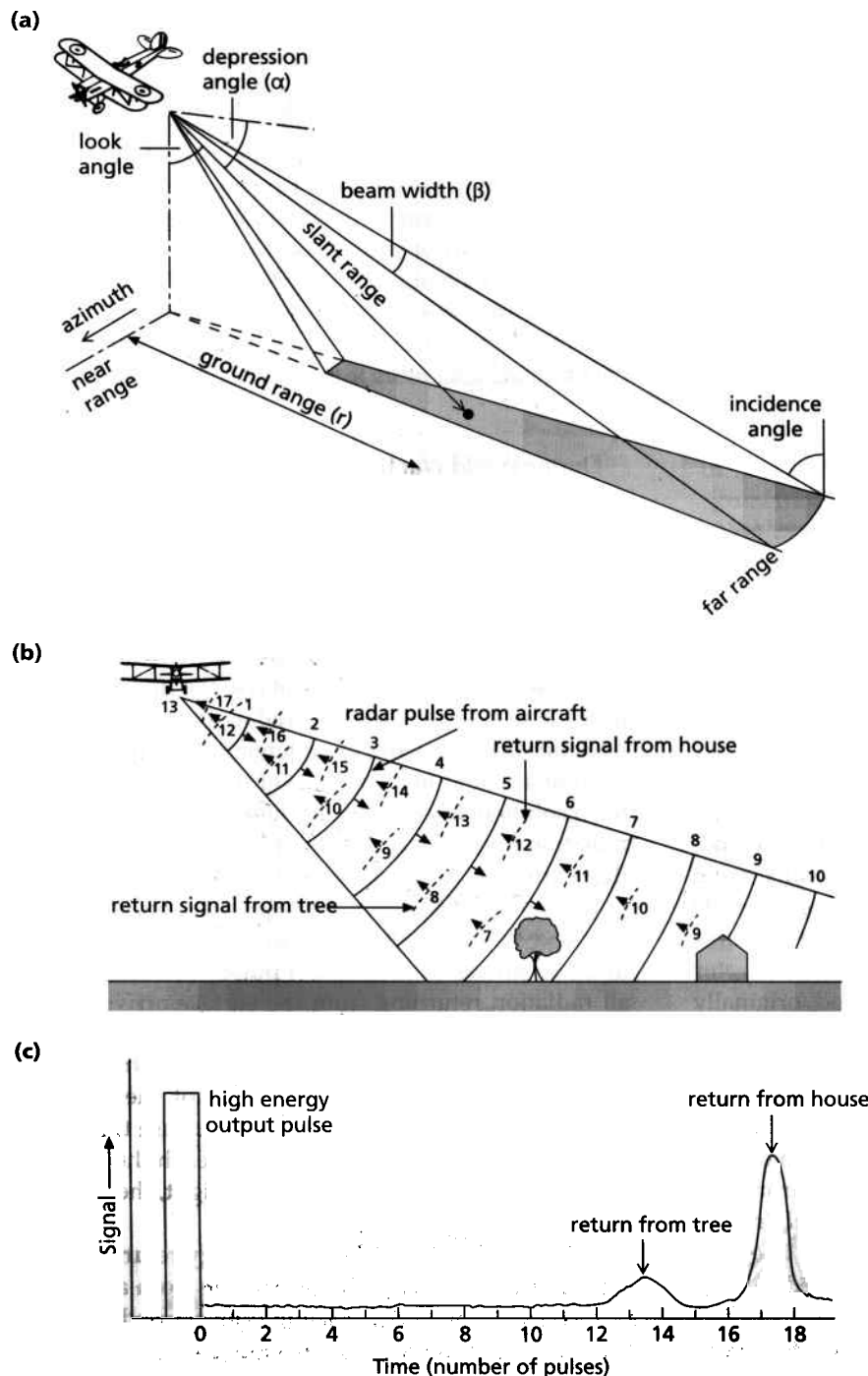


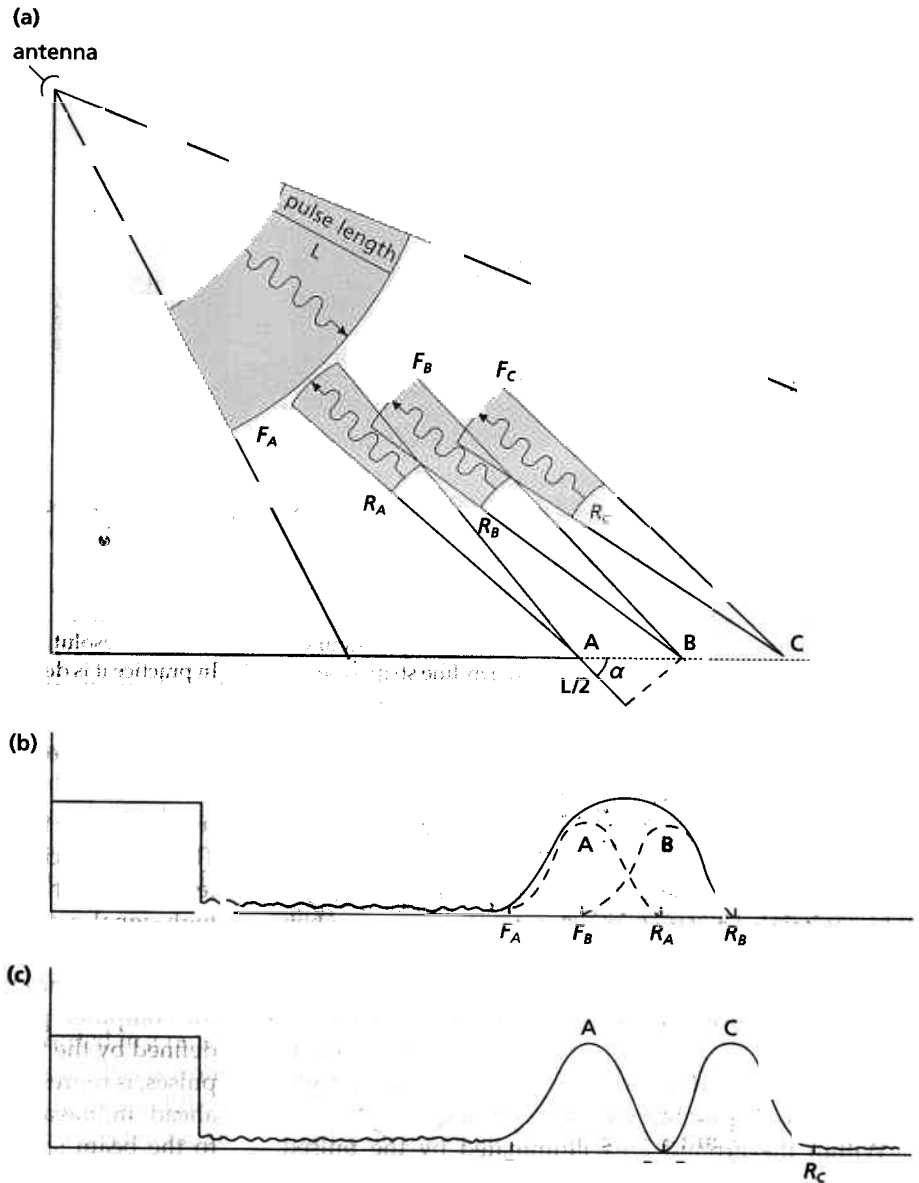
Fig. 3.14 (a) The geometry involved in sideways-looking radar governs many of the intrinsic properties of radar images. The depression angle (α) is the angle between horizontal and a radar ray path. The look angle, sometimes quoted instead, is simply the angle between vertical and a ray path ($90^\circ - \alpha$). The incidence angle is the angle between an incident radar ray and a line at right angles to the surface. For a horizontal surface it is the same as the look angle, but varies with the slope of the surface. The beam width (angle β) determines how the illumination of the surface spreads out from near to far range. The slant range is the direct distance from the antenna to object, and is related to the true or ground range by the depression angle (Equation 3.7). (b) The solid curves on this cartoon are spherical radar wavefronts emitted by the antenna, and their numbers indicate the time since they were emitted. The dashed curves are radar wavefronts back-scattered from the house and tree. They have the same number convention indicating time. The wavefront from the house which has just been received at the antenna has travelled for 17 time units, that from the tree for 13. The graph (c) shows how radar energy returned from the house and tree is recorded at the antenna.

single pulse and its return from objects on the ground (Fig. 3.14b). With the next pulse the platform has moved forwards, to acquire another line of data. At its most simple the image is built up by modulating the intensity of a single line cathode-ray tube with the energy-time signal. Exposing film to this at a rate proportional to the platform speed builds up a strip-like image.

Resolution on a radar image depends on two parameters: pulse length and beam width. The length of each pulse (time τ , distance $c\tau$) controls the resolution

in the range direction. To distinguish two objects on the ground, the returned signals from each object for the whole pulse must reach the antenna at different times. Any overlap between the two sets of signals means that the objects merge into one on the image. As Fig. 3.15 shows, the minimum separation between the objects along the slant range must be half the pulse length, or $c\tau/2$. This limiting distance, the slant-range resolution, is independent of distance from the platform. The corresponding distance on the ground, the ground-range

Fig. 3.15 In (a), the surface is illuminated by pulses of radar from the antenna (grey) of a fixed duration τ . The pulse length L is equal to $c\tau$. The fronts of back-scattered returns of a pulse from objects at A and B are at F_A and F_B , and the rears at R_A and R_B . The signals that these returned pulses generate at the antenna (b) overlap so that they are just indistinguishable, and form a single broad peak. Pulses returned from objects with a closer spacing than A and B would overlap even more. At spacings greater than AB, however, return signals are separate (c) and the two objects, such as A and C, can be distinguished. The limiting slant range corresponding to the separation between A and B is half the pulse length L . This is the slant-range resolution. In the ground range this resolution becomes $L/2 \cos \alpha$.



resolution (R_r), however, varies with the depression angle (α), so that:

$$R_r = (c\tau \cos \alpha) / 2 \quad (3.8)$$

and it increases away from the platform. However, it is theoretically independent of the altitude of the platform.

The width of the radar beam is expressed by an angle (β) (Fig. 3.14a), and is directly proportional to the radar wavelength (λ) and inversely proportional to the length of the antenna (L). The width of the line on the ground illuminated by each pulse depends on the distance from the platform. As distance increases, so the beam fans out. The resolution parallel to the ground track of the platform, the azimuth resolution (R_a) is the width illuminated by each pulse. Azimuth resolution deteriorates in direct proportion to the ground range (r)

$$R_a = r\beta = (\lambda/L)(c\tau \cos \alpha) / 2 \quad (3.9)$$

Each line on an image therefore represents data that become increasingly degraded away from the platform. The azimuth resolution also deteriorates as the altitude of the platform becomes greater. So a satellite radar system based on sensors developed for aircraft would have equally good range resolution, provided the power was high enough, but hopeless azimuth resolution. There are, however, ingenious ways of avoiding this limitation.

The most obvious way to decrease the beam width is by extending the antenna. From an altitude of 100 km an azimuth resolution of 100 m requires a beam width of $1 \mu\text{rad}$. For 25 cm wavelength radar, substituting in Equation (3.9), shows that an antenna 250 m long will be required. Clearly, there are physical limits using this

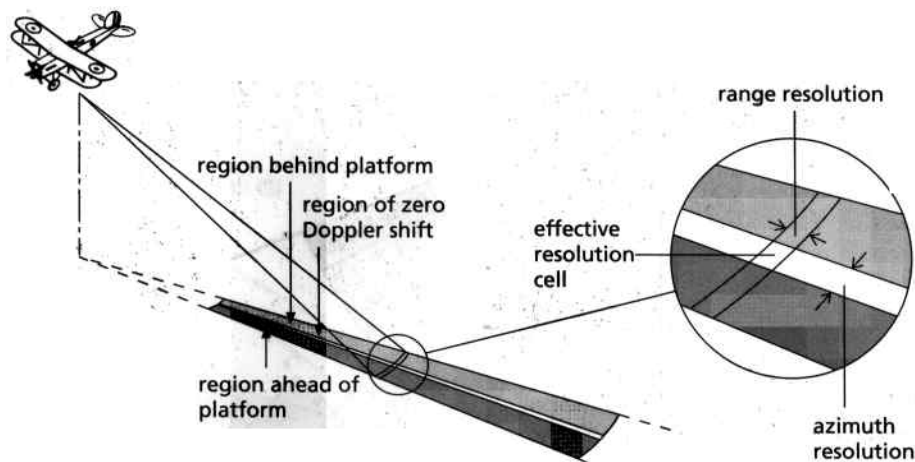


Fig. 3.16 In a synthetic aperture radar system interference occurs between a reference signal and the increased or decreased frequencies of radar waves returning from illuminated areas ahead of and behind the platform. Waves returning from a narrow strip at right angles to the flight path do not suffer a Doppler shift, and no interference occurs. This allows returns at any instant from the in-line strip to be discriminated from those from the rest of the illuminated surface during signal processing. This makes possible a much finer resolution in the azimuth direction than could be

solution. Radars for which beamwidth is controlled by the length of their antennae are known as real-aperture systems. They have simple designs and recording systems, as outlined earlier, but they are useful only when mounted in aircraft. Real-aperture radar is now obsolete. More complex systems use the coherent nature of artificial microwave radiation to simulate an extremely long antenna. They are called synthetic-aperture radar (SAR) systems, and are the only useful means for high-altitude and satellite-radar remote sensing.

Within the ground area illuminated by the pulsed radar beam (Fig. 3.16), parts of the surface are ahead of the platform, parts behind, and only a narrow strip, just as wide as the antenna itself, is in line at any instant. The returns from the areas ahead and behind are subject to a Doppler shift, just as sound would be. Imagine a train with its horn sounding. Sound echoed from ahead would have a higher pitch to an observer on the train than would that echoed from behind. Only echoes from the side would sound the same as the horn. The Doppler effect has altered the wavelength of the echoed sound. Exactly the same thing happens to radiation, including radar, although the shift is very small. Because radar is coherent these small differences can be detected electronically.

A SAR system records the time, the energy and the frequency of radar returns from the surface, with different frequencies signalling the position of the object responsible for a frequency shift relative to the motion of the platform. The frequency is measured by allowing the returns to interfere with a reference signal with a

frequency that is the same as that of the radar pulse itself. Those frequencies affected by Doppler shift produce interference patterns. A constructive interference produces a high signal, a destructive interference gives a low signal. These can be recorded optically as bright and dark spots arranged in lines parallel to flight direction, or digitally for computer processing. Each object on the ground, defined by the time of the response for successive radar pulses, is represented by several spots, from when it was ahead, in line and behind the moving platform. Owing to the beam spread, the further an object is from the platform the more times the electronics 'look at' it. That transforms the moving real antenna into a much longer, synthetic antenna, thereby improving the azimuth resolution. Because the spread of the beam illuminates far-range objects for longer than those at near range, azimuth resolution is independent of range. That electronic trick makes SAR imaging possible from satellites, subject only to the constraints of weight, power requirements and the rate of data transmission. The last is enormous by comparison with data from optical systems, because of the great complexity of 'raw' SAR data.

Recording SAR signals on film produces a one-dimensional interference, in the form of a radar hologram (Fig. 3.17a). By itself it is meaningless, but when illuminated by a laser it is optically correlated to produce a radar image (Fig. 3.17b). Digital SAR signals can be unpacked by very powerful software to produce a digital image in raster format, which suffers none of the degradation inherent in optical correlation.

Recording SAR signals on film produces a one-dimensional interference, in the form of a radar hologram (Fig. 3.17a). By itself it is meaningless, but when illuminated by a laser it is optically correlated to produce a radar image (Fig. 3.17b). Digital SAR signals can be unpacked by very powerful software to produce a digital image in raster format, which suffers none of the degradation inherent in optical correlation.

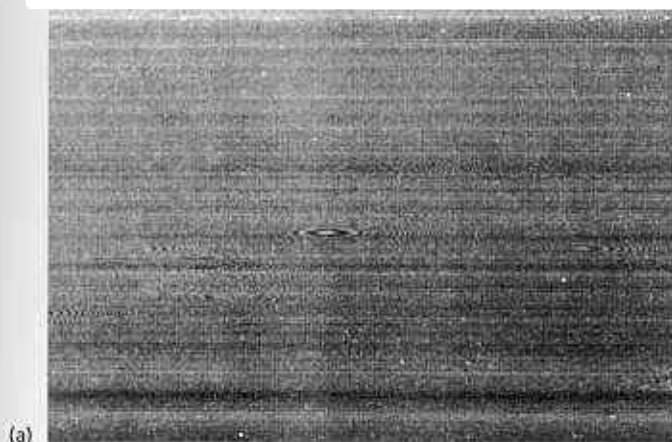


Fig. 3.17 The raw data from a synthetic-aperture radar system is expressed on film as a hologram (a), which cannot be interpreted until it has been optically correlated. The patterns on the film represent interference between the Doppler-shifted frequencies of back-scattered radar waves and a reference



frequency generated by the system. Projection of a laser through a SAR hologram using a complex lens system optically correlates the interference patterns on the hologram to produce an image of the surface (b). Courtesy of Eric Kasischke, Environmental Research Institute of Michigan.

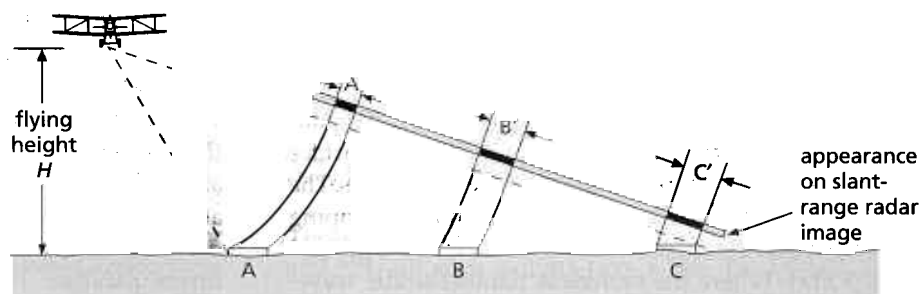
Radar transmission can be in different modes of polarization, with vibrations restricted to a vertical or horizontal plane, as can receiving the returns. Such configurations enable the rotations induced by reflections from different types of surface materials to be analysed. This and other surface-dependent features of radar images form the central part of Chapter 7. We turn now to the appearance of radar images.

Because radar looks sideways, and as radar images represent energy and time, it produces unusual geometric features as well as those resulting from platform instability and Earth rotation. Times recorded in a radar image represent direct or slant distances of objects from the platform, and each surface object subtends a different angle to the platform. As a result, uncorrected radar images vary in scale along range. They are compressed at near range and only approach true scale at far range, the opposite of line-scan images. This results from the different incidence angles of radar wavefronts at different ranges (Fig. 3.18) and is an inherent distortion

in slant-range images. A hyperbolic transformation of the data down range converts them to a nearly uniform scale, which equates that of a flat surface, thereby producing a ground-range image.

Surfaces with high relief produce a strange effect. The top of a mountain may be nearer to the platform than its base. Because radar produces time images, the top appears closer to the platform than the base (Fig. 3.19). The mountain appears to 'lean' towards the platform, giving a phenomenon termed layover. Many radar images of mountainous terrain show arrays of layovers (Fig. 3.20), which resemble the flatirons normally associated with dipping strata (Chapter 4). They have no relationship to dip and are purely artificial. The layover effect depends partly on the relief, but mainly on the depression angle. For high depression angles—when the surface is more nearly below the platform—the effect is worst. With a shallow depression angle the effect is noticeable only for the strongest relief. The effect at its most bizarre can display a mountain overlaying a river or glacier (Fig. 3.22b).

Fig. 3.18 In a slant-range image the scale down-range increases because the depression angles of spherical wavefronts which meet the surface decrease away from the antenna. The apparent distances between objects and between parts of the same object are measured in time. The time taken to illuminate A is shorter than for B and C. Likewise it takes longer for a wave to move from B to C than from A to B. The equal distances on the ground increase away from the platform on the slant-range image.



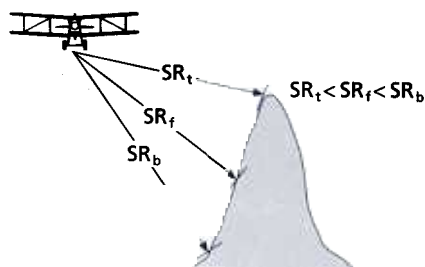


Fig. 3.19 Because the whole of its slope that faces the platform is at a steeper angle than a radar wavefront, the top of the mountain is illuminated by the wavefront before its flank and base. This is translated into slant ranges— SR_t , SR_f and SR_b —which produce an image where the top is nearer to the platform than the base or flank. The mountain appears on the image to lean towards the platform.

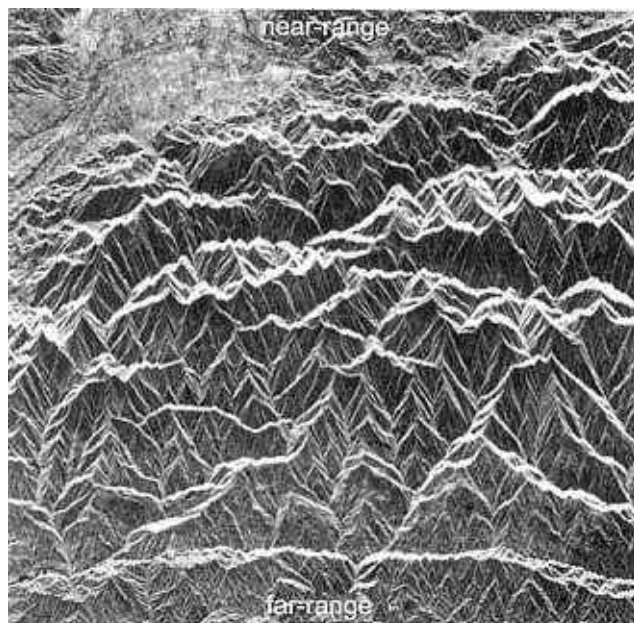


Fig. 3.20 Layover dominates both the near- and far-range parts of this radar image of a mountainous area in Japan. It takes the form of arrays of triangular features that look like flatirons, not to be confused with real 'flatirons' which are so characteristic of eroded dipping strata on aerial photographs and other images in the visible and infrared. Note how some of the mountains appear to 'lie over' drainage courses. Courtesy NASDA, Japan.

Related to layover is another effect typical of radar images: foreshortening. This occurs where the surface slope is less steep than that of the radar wavefront. The pulse reaches the base before the top, so that the sloping surface appears shorter on the image than on the ground (Fig. 3.21a). Where the surface is parallel to the wavefront all points on it are exactly the same slant distance from the platform, so that all fall together on the image as

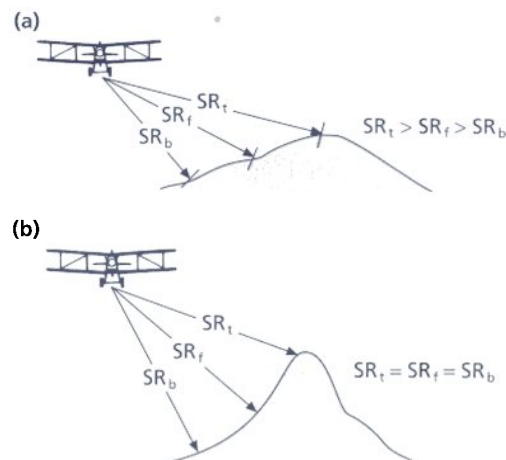


Fig. 3.21 Where the surface slope is at a lower angle than radar wavefronts (a) the slant ranges of top, flank and base of a mountain are less than their true ranges on the ground. The result is that the slope facing the platform is compressed or foreshortened. Sometimes the surface has approximately the same angle as a wavefront (b). In this case the slant ranges for all parts of the surface are the same and the whole slope is compressed to a single line.

a single point (Fig. 3.21b). Examples of both phenomena can be seen in Figs 3.20 and 3.34(a).

The most striking feature on radar images of variable relief is their sidelit character (Figs 3.20 and 3.34a). Slopes facing the platform are most strongly illuminated and reflect most of the energy back to the antenna, so showing up as bright. Slopes facing away receive least energy and appear darker. Where such slopes are steeper than the radar ray paths, then they are in radar shadow, receive no energy and appear totally black (Fig. 3.22). The farther the range, the lower the depression angle and the lower the relief needs to be to produce a radar shadow.

Where depression angle varies greatly from near to far range, as in all images acquired from aircraft, layover, foreshortening and shadowing vary markedly in the range direction (Figs 3.20 and 3.34a). This can be so grotesque in areas of only moderate relief that such images are of little use for detailed interpretation. One escape from this problem is to mosaic strips with the same depression angle from overlapping swaths. A better solution is to make images from such an altitude that only a small range of depression angles covers a very broad swath on the ground. The best images are those from orbit. Despite their irritations, these peculiarities inherent in radar image have one very useful advantage. Differential relief displacements caused by different depression angles and look directions cause image parallax. That, as in overlapping aerial photographs, enables stereoscopic viewing (Chapter 7 and Appendix A).

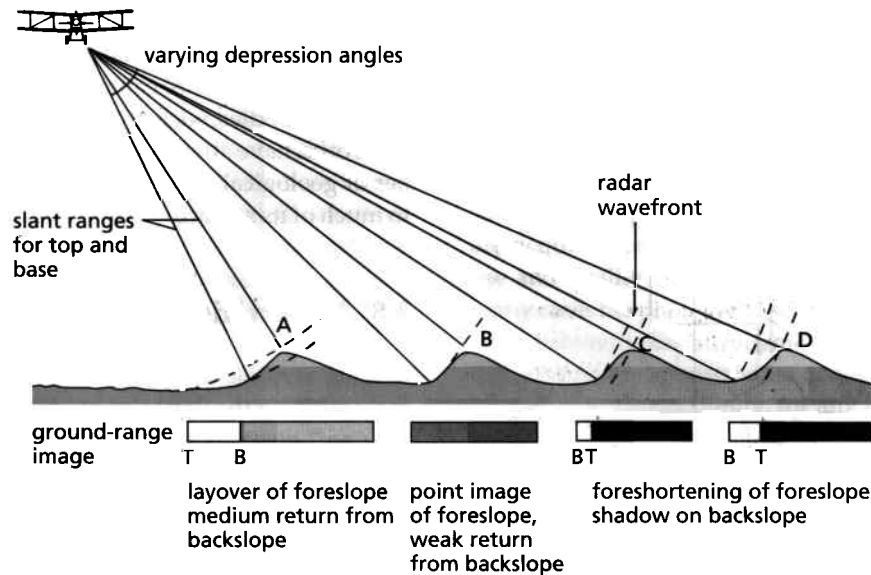


Fig. 3.22 This diagram summarizes the peculiarities of radar image geometry by showing the effects on mountains with the same shape at different ranges. Layover dominates slopes facing the platform at near range, and foreshortening falls from a maximum at medium depression angles to less significance in the far range. Wherever they are, slopes steeper than wavefronts always suffer layover, whereas shallower slopes are always foreshortened to some extent. Surfaces that

slope away from the platform receive less energy and so always appear darker on a radar image. If their slope is greater than the depression angle they receive no energy and are in total shadow. As depression angle decreases towards the far range, this part of an image is more likely to contain shadows than the near range. Figure 3.21 shows examples of layover and foreshortening from near to far range.

3.6 Imaging spectrometers

The spectral curves shown in Chapter 1 are products of spectroradiometers. They direct radiant energy from a surface through a diffraction grating to spread the radiation into its constituent spectrum. Detectors appropriate for different wavelength ranges scan the spectrum, recording radiances for very narrow wavebands, so constructing accurate spectral curves. Although useful in establishing the spectral characteristics of materials, such instruments are not appropriate for mapping purposes. Although it is possible to carry spectroradiometers on remote-sensing platforms, the various factors discussed in Section 3.3 around Equation (3.5) limited them until recently to only a few narrow wavebands in producing profiles along flight lines. Carefully chosen bands could allow important information on the distribution of different minerals to be produced, but such systems were unwieldy and limited in the data that they produced.

The development of CCDs sensitive across the whole reflected region of the spectrum, and the ability to miniaturize them in large arrays, allows much more flexible systems. Exposing such an array, with different CCD elements 'tuned' through appropriate optics to adjacent narrow wavebands, produces detailed spectra for small areas along a profile. Plate 3.2 shows an example that has been enhanced digitally in order to emphasize spectral features.

The most revolutionary outcome of CCD arrays is the development of imaging spectrometers. Instead of producing a few broad spectral bands for each pixel in an image, they capture hundreds of very narrow bands in image form. One method uses a long linear array of CCDs, very similar to that in the pushbroom device described in Section 3.4, but having CCDs sensitive to different parts of the spectrum arranged along it. A mirror focuses radiant energy through a diffraction grating, so that the resulting spectrum spreads along the CCD array. This combination of line-scanning with a long CCD array is a whiskbroom imaging spectrometer. By combining several linear CCD arrays, each tuned to adjacent narrow spectral bands, in an area array n bands long by m elements wide, an even simpler pushbroom imaging spectrometer is possible. This operates in exactly the same general fashion as the system shown in Fig. 3.13.

The many spectral channels, combined with millions of pixels in images from imaging spectrometers, pose problems of data storage and means of analysing the data. Some of these techniques are covered in Chapter 5.

3.7 Gamma-ray spectrometers

The naturally occurring unstable isotopes ^{40}K , ^{232}Th , ^{235}U and ^{238}U all emit gamma-rays when they decay. Each isotope (more precisely, various daughter isotopes in their decay schemes) emits a range with discrete energy

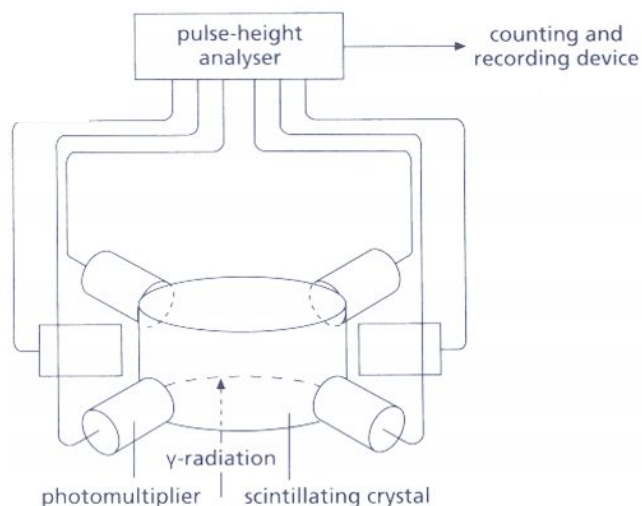


Fig. 3.23 A gamma-ray spectrometer consists of a crystal in which gamma-ray photons induce scintillation. The light emitted is intensified by photomultipliers and measured by a pulse-height analyser. The amount of light produced by each photon is proportional to the gamma-ray energy level. The number of light pulses with a particular level are counted over a fixed time, and represent the intensity of gamma-rays with the respective energy level.

levels that a gamma-ray spectrometer can detect and measure. Such instruments depend on the emission of bursts of light by certain crystals, such as sodium iodide doped with thallium, when they absorb gamma-rays. The intensity of the light is proportional to the energy and inversely proportional to the wavelength of the gamma-ray photons, whereas the frequency of light pulses is a measure of the amount of radioactivity. By filtering the discrete energy levels and measuring pulse frequency for narrow energy bands it is possible to estimate the proportion of each isotope in the surface which emits the gamma-rays.

In practice a gamma-ray spectrometer comprises a very large crystal surrounded by photomultiplier tubes and appropriate shields, which amplify the bursts of light and block out cosmic rays, respectively (Fig. 3.23). Measurements are made with respect to another crystal which receives continuous gamma-rays from a standard radioactive source.

When first devised, this technique was capable of accurate analysis only under laboratory conditions, or for very crude analogue measurements from the air. Improvements in electronics, photomultipliers and the size of suitable crystals now mean that quantitative airborne data are possible. Systematic overflights in a grid pattern, together with reformatting and interpolation of data to a raster produces images that represent the apparent uranium, thorium and potassium contents of surface materials (Chapter 8).

The main hindrances to gamma-ray imaging result from the way in which the atmosphere and vegetation

cover absorb the radiation, thereby attenuating it. For terrestrial uses only aerial surveys, flown at ground-hugging altitudes, are capable of producing useful data. For airless bodies, such as the Moon, gamma-ray surveys from orbit have allowed the extrapolation of a tiny number of geological observations by the Apollo astronauts to much of its surface.

3.8 A short history of remote sensing

Remote sensing depends on the ability to get sufficiently far from the Earth's surface to obtain a worthwhile view and the invention of means of capturing an image. Although the first staffed balloon flight was in 1783, daguerreotype photography appeared only in 1839. It was not until 20 years later still that these two achievements were combined, when Gaspard Tournachon ascended in a balloon and photographed a village near Paris.

Military applications have always dominated the development of remote sensing since then. During the Civil War (1861–65) in the USA, the Union forces photographed Confederate defensive positions from captive balloons, and using remotely operated cameras flown on kites. Attempts were even made in Germany to use delayed-action miniature cameras carried by pigeons. During World War I (1914–18), aerial photographic reconnaissance using powered aircraft systematized aerial photography. That enabled its first significant uses in cartography, forestry and geology during the 1920s and 1930s. The first major geological applications were for oil exploration in what was then Persia, by the Anglo-Persian Oil Company. Aerial photo-reconnaissance dominated planning and execution of many operations during World War II. However, three new developments were achieved under the spur of military necessity. Radar was developed, together with VNIR-sensitive film for camouflage detection and trials were made of thermal-infrared sensing devices.

During the 1950s, colour infrared photography found uses in vegetation studies, side-looking airborne radar was refined and synthetic-aperture radar became possible with the development of a successful optical processor. In the same decade photographic technology and interpretation skills continued to develop, with the clandestine flights of US Air Force spy planes such as the supersonic U2 over Soviet and other territory.

Since the early 1960s there have been large numbers of military surveillance satellites in orbit. Many of the earlier ones recorded images on film, which was recovered on re-entry of the platform, ejected in canisters to be scooped up by recovery aircraft or located electronically on the ground. Increasingly, however, intelligence gathering has moved towards electronic means using line-scanners and now pushbrooms, depending on ever increasing spectral

sensitivity of the solid-state detectors. These devices were at first military secrets, but as pressures grew for civil and commercial applications, the technologies were progressively declassified in the USA. In retrospect, it seems clear that that secrecy lay not in the technological principles involved, but the limits to which they could be pushed and the precise areas on which attention was directed. Various lines of indirect evidence suggest that satellite imagery with resolution better than 1 m has been available for two decades. The present estimate of 10–15 cm visible–VNIR resolution from the latest US ‘Keyhole’ satellite series (KH-12) may indeed have been superseded. Ultrafine resolution radar and thermal imagery is also indicated by indirect evidence. Probable global coverage and certainly global potential is possessed by USA and the former Soviet intelligence communities, but declassification for civilian applications is likely to be slow. The reason for this is likely to be a mixture of diplomatic nicety and the advantage of potential opponents not knowing precisely what can be hidden and what can be seen. In 1996 the US Government authorized phased public release of satellite photography from the Corona series of intelligence satellites through the US Geological Survey, but commercial systems now offer better resolution (around 1 m).

In the next three sections remotely sensed data that can be acquired openly, albeit at some cost, and which are of use in geology are described.

3.9 Airborne data

Remote sensing from aircraft has several advantages over orbital remote sensing, but there are also disadvantages. It is easier to change sensors to suit a particular study, and to record at any time of day. The flying height can be varied, from less than 100 m above ground to 30 km to give flexibility of scale, and to reduce various atmospheric effects. Major disadvantages are that an aircraft is subject to various perturbations which can cause distortions in the images, and that it is sometimes not possible for the aircraft to fly in sensitive areas.

3.9.1 Aerial photographs

Aerial photographs are the most easily available of all remotely sensed data, and have been acquired at various scales for much of the land area of the Earth. Their first widespread use was in World War I as a source of military intelligence. Soon afterwards oil companies began to use them in geological mapping. The basic theory and practice of making cartographic measurements from stereoscopic pairs of overlapping vertical aerial photographs were developed in the interwar years. Major advances in film and camera technology took place during World War II,

again at the demand of the military. Since then their acquisition has become a matter of routine for a wide variety of agencies, following the transfer of technology from military to civilian applications.

Aerial photographic surveys are commissioned by both national and local government agencies and by consulting and exploration companies. In the first case they are acquired primarily for monitoring changes in land use and quality, and for updating topographic maps. In many cases the photographic campaigns are co-ordinated with conventional surveying, when requirements for different scales sponsor data acquisition from a variety of altitudes. Consequently, for any area of interest to the geologist, there may be photographs with different fields of view, scales and resolutions for a number of different years and seasons.

The basic form of an aerial survey consists of a number of approximately parallel flight lines, along which exposures are timed to give at least 60% overlap between successive photographs, and the spacing of which is such that adjacent lines of images have sufficient sidelap to ensure complete coverage (Fig. 3.24). More often than not, an aircraft cannot maintain a perfect course, because of crosswinds and turbulence. If no correction

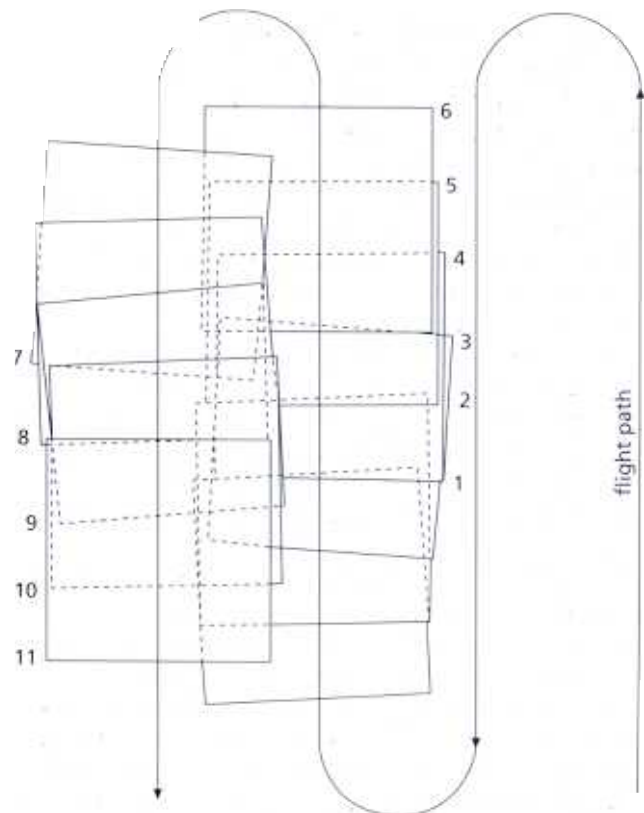


Fig. 3.24 The aircraft that took the individual photographs outlined here had to change its heading to follow the planned flight path because of wind effects. Many of the frames are therefore twisted out of alignment.

is made for crosswinds the photograph edges remain parallel to the intended flight line, but they drift away from that line (Fig. 3.24). Correction normally means turning the aircraft's nose into the wind, which causes it to crab, and if the camera is not reoriented the photograph edges cease to parallel the flight line, although their principal points will lie along it (Fig. 3.24). Usually these deviations are not so severe as to affect seriously the usefulness of the photography. Turbulence does cause problems, however. The aircraft may rise and fall, changing the scale of each image. It may pitch or roll, imparting a tilt to each image which produces unreal slopes in the stereomodel (Fig. 3.7). Finally, the ground speed may change, thereby altering the overlap between adjacent images. If the overlap is less than 50% there will be strips across the images for which stereoviewing is impossible. If it is greater than 70% the stereomodel will show insufficient relief for proper interpretation.

3.9.2 Other airborne data

Airborne scanners and pushbroom systems are flown commercially by a number of private companies and by national military and civil organizations. Some of them have been used in campaigns to give scientists

experience in how to use and interpret similar data from future satellite systems. Table 3.2 shows some of the specifications for two such airborne systems operating in the reflected and emitted regions, each of which has produced large volumes of data both for public use and confidential exploration. The most interesting of these is the Geoscan, operated by Carr-Boyd Pty of Australia, which incorporates spectral coverage from both the reflected and emitted regions. The wavebands in the reflected region are sufficiently narrow and positioned to enable discrimination of specific mineralogical features in spectra. The bands in the emitted region are, likewise, targeted on molecular features of emission spectra. It combines fine spectral resolution within a strategically placed but limited number of wavebands that keeps data volume within manageable limits without sacrificing the system's potential for detecting and discriminating different surface materials.

More innovative airborne systems include the NASA imaging spectrometers. The first of these (1983) was the Airborne Imaging Spectrometer (AIS), which used a pushbroom, area-array detector (Section 3.6). This was superseded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) based on the whisk-broom principle. Together with a number of similar,

Table 3.2 Spectral range of three commercial airborne multispectral imaging systems

| Daedalus AADS 1268 IFOV 2.5 or 1.25 mrad | | | Geoscan AMSS Mk II IFOV 2.1 mrad | | | TIMS IFOV (2.5 mrad) | | |
|---|---------------------------------------|-----------------------------|-------------------------------------|---------------------------------------|-----------------------------|-------------------------|---------------------------------------|-----------------------------|
| Band number | Central wave-length (μm) | Bandwidth (μm) | Band number | Central wave-length (μm) | Bandwidth (μm) | Band number | Central wave-length (μm) | Bandwidth (μm) |
| 1 | 0.435 | 0.030 | 1 | 0.522 | 0.042 | 1 | 8.4 | 0.4 |
| 2 | 0.485 | 0.070 | 2 | 0.583 | 0.067 | 2 | 8.8 | 0.4 |
| 3 | 0.560 | 0.080 | 3 | 0.646 | 0.071 | 3 | 9.2 | 0.4 |
| 4 | 0.610 | 0.020 | 4 | 0.693 | 0.024 | 4 | 9.8 | 0.8 |
| 5 | 0.660 | 0.060 | 5 | 0.717 | 0.024 | 5 | 10.7 | 1.0 |
| 6 | 0.720 | 0.060 | 6 | 0.740 | 0.023 | 6 | 11.7 | 1.0 |
| 7 | 0.830 | 0.140 | 7 | 0.830 | 0.022 | | | |
| 8 | 0.980 | 0.140 | 8 | 0.873 | 0.022 | | | |
| 9 | 1.650 | 0.200 | 9 | 0.915 | 0.021 | | | |
| 10 | 2.150 | 0.270 | 10 | 0.955 | 0.020 | | | |
| 11 | 10.500 | 4.000 | 11 | 2.044 | 0.044 | | | |
| | | | 12 | 2.088 | 0.044 | | | |
| | | | 13 | 2.136 | 0.044 | | | |
| | | | | 2.176 | 0.044 | | | |
| | | | 15 | 2.220 | 0.044 | | | |
| | | | 16 | 2.264 | 0.044 | | | |
| | | | 17 | 2.308 | 0.044 | | | |
| | | | 18 | 2.352 | 0.044 | | | |
| | | | 19 | 8.640 | 0.530 | | | |
| | | | 20 | 9.170 | 0.530 | | | |
| | | | 21 | 9.700 | 0.530 | | | |
| | | | 22 | 10.220 | 0.530 | | | |
| | | | 23 | 10.750 | 0.530 | | | |
| | | | 24 | 11.280 | 0.530 | | | |

Table 3.3 Specifications of two imaging spectrometers

| | AVIRIS | HIRIS |
|--------------------|-------------------------|-------------------------|
| Platform | Aircraft | Satellite |
| IFOV (mrad) | 2.5 | 8.8 (30 m pixels) |
| Swath (pixels) | 614 | 800 |
| Spectral range | 0.40–2.45 μm | 0.40–2.45 μm |
| Number of channels | 224 | 192 |
| Bandwidth (nm) | 10 | 10–20 |

but less-widely deployed imaging spectrometers, these devices are precursors to the High-Resolution Imaging Spectrometer (HIRIS) intended to fly in orbit in the early 2000s (Section 3.13). Details of the specifications of AVIRIS and HIRIS are given in Table 3.3.

Airborne imaging radar is now used routinely by oil and mineral exploration companies over terrains that are often cloud covered. Country-wide SAR surveys by governments are becoming widespread, the most notable being the Brazilian RADAM project in the 1970s and the ongoing US Geological Survey programme for the whole USA. Extensive multipolarized and multispectral radar surveys have been flown by NASA, principally in North America, as forerunners of Earth orbital systems and interplanetary probes. Examples of airborne SAR images are given in Chapter 7.

3.10 Basic characteristics of orbiting satellites

The orbit of a satellite is elliptical in shape, but remote sensing satellites are usually put in orbits that are very close approximations to a circle. The laws of gravity dictate that a satellite in a higher orbit travels more slowly than one in a lower orbit, and because higher orbits are longer than lower ones this means that a high orbiting satellite takes a good deal longer to circle the Earth than a low orbiting satellite.

For a satellite in a circular orbit of radius R_o from the centre of the Earth its velocity v is given by:

$$v = (GM/R_o)^{1/2} \quad (3.10)$$

where G is the universal gravitational constant ($6.67 \times 10^{-11} \text{ Nm}^2 \text{ kg}^{-2}$) and M is the mass of the Earth ($5.98 \times 10^{24} \text{ kg}$).

The time taken to complete an orbit, the orbital period, P is given by:

$$P = 2\pi R_o / v \quad (3.11)$$

The radius of the Earth (R_e) is 6371 km, and so the height (h) of the orbit about the Earth is:

$$h = R_o - R_e \quad (3.12)$$

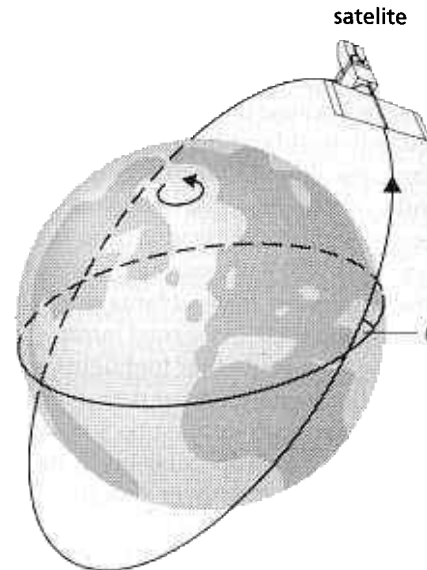


Fig. 3.25 The orbital inclination of a satellite is given by an angle θ between the plane of the orbit and that of the Equator. If the satellite passes directly over the poles then the inclination is 90° . Conventionally, for orbits that take the satellite round the globe in the same direction as the Earth's rotation, θ is less than 90° ; if it travels in the opposite direction then θ is greater than 90° .

Below 180 km the Earth's atmosphere is too dense for satellites to orbit without uncontrollable frictional heating. Above 180 km there is still a small atmospheric drag on a satellite, causing its orbit to gradually spiral downward until eventually it reaches thicker atmosphere and burns up. Satellites are launched into such low orbits, but they do not last long. Above a few hundred kilometres there is so little atmospheric drag that a satellite will remain in high orbit indefinitely.

The plane of an orbit must always pass through the Earth's centre but can be in any orientation. If its orbit is inclined at more than 45° to the equatorial plane then a satellite is in polar orbit. One that is less steeply inclined is called an equatorial orbit (Fig. 3.25). Various forces other than atmospheric drag can perturb an orbit, such as the gravitational attraction of the Sun and Moon. Jets of gas maintain the orbit and also control the spin of the satellite.

An orbiting satellite normally changes its position in the sky continuously. However, if a satellite has an orbital period exactly the same as that in which the Earth rotates and orbits in the same direction as the Earth rotates, it is in a geosynchronous orbit. To do this its orbital altitude must be 35 786 km. If a satellite is in a circular geosynchronous orbit with zero inclination (above the Equator) then it appears to remain stationary over the same point on the ground, and is in a geostationary orbit. Geostationary orbits are useful for communications and meteorological satellites, which need to be sited sufficiently far

from the globe that they can monitor almost a whole hemisphere on a nearly continuous basis. The altitude of geostationary satellites is so great that the ground resolution is low because the smallest IFOV of sensors in the reflected region is about $20\ \mu\text{rad}$, which at an altitude of 35 786 km gives about 720 m on the ground.

It is useful to obtain repeated images at the same time of day, to ensure similar illumination conditions. The time can be selected to avoid early morning mists or afternoon cloud in the tropics, to take advantage of the diurnal heating and cooling cycle in thermal infrared studies, or to ensure a low Sun angle that highlights topographic features. To do this a satellite must pass over all points at the same local solar time—it must be Sun-synchronous. A satellite placed in a steep polar orbit, the motion of which carries it westwards over the ground at a rate comparable with the Earth's rotation, achieves this condition. Most Sun-synchronous polar orbiters pass from north to south over the sunlit hemisphere and return from south to north over the night-time hemisphere. Because the Earth rotates beneath the orbital plane, successive orbits cover different ground tracks. As the orbital period is much less than 1 day, images of several ground tracks can be acquired within 24 h. Most Sun-synchronous remote-sensing satellites are placed in orbits with altitudes between 600 and 1000 km. At 800 km, a sensor with an instantaneous field of view of $20\ \mu\text{rad}$ would have a ground resolution cell of 16 m, compared with the 720 m achieved from geosynchronous orbit.

The type of ground coverage offered by a Sun-synchronous satellite is shown in Fig. 3.26. Unless the swath is very wide, there is a gap between the swaths imaged in successive orbits. Part of the design of orbiting systems is to calculate an orbital geometry that ensures that the gaps are filled as efficiently as possible. High-

resolution systems with swath widths of 100–200 km can completely image the globe within about 20 days, whereas low-resolution systems, with swath widths of over 1000 km, can image the globe in the course of a single day.

One disadvantage of Sun-synchronous orbits is that it is not always possible to transmit an image to the ground as soon as it is gathered, in real time. To send a signal directly to the ground means that the satellite has to be above the horizon as seen from a ground receiving station. At typical orbit heights, the satellite has to be within about 3000 km of the station. If there is no ground receiving station in line of sight from the satellite then the image is either transmitted to the ground via a relay satellite in a higher, geosynchronous orbit or is recorded on-board the satellite to be transmitted to a ground station when it comes within range. Tape recorders use power and make the satellite heavier to launch, so not all remote-sensing satellites carry them.

In the case of radar imagery a Sun-synchronous orbit is unnecessary, because illumination is active. However, only polar orbits allow full global coverage.

Spacecraft that carry personnel are much heavier than unstaffed ones, because they must include life-support systems. This means that it is difficult to lift them into high orbits and expensive to put them in polar orbit. Equally important, it is difficult to recover personnel safely. Usually their orbital inclination is about the same as the latitude of the launch site. Most staffed missions are of short duration, which makes them unsuitable for monitoring at a fixed solar time and incapable of acquiring comprehensive cover even within the limits of their orbital inclination. Staffed missions, such as the Space Shuttle, tend to be used as test-beds for experimental systems. For example, spaceborne imaging radar was advanced considerably by Shuttle experiments.

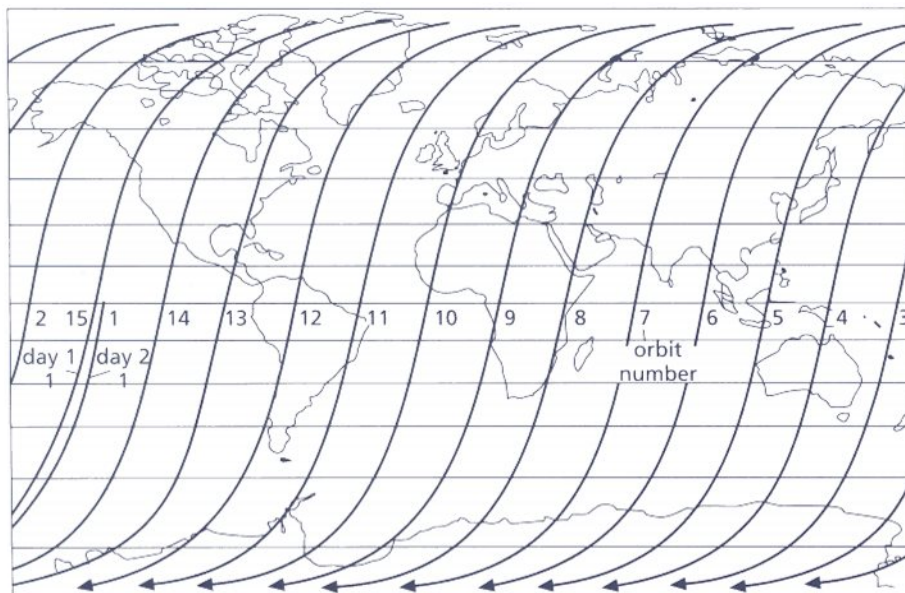


Fig. 3.26 The groundtrack of a Sun-synchronous satellite (in this case the first of the Landsat series) for a single day. In fact the local solar time varies from 1030 hours at 60°N to 0830 hours at 60°S , and a truly Sun-synchronous orbit is impossible.

3.11 Data from staffed spacecraft

Staffed satellites have provided platforms for a wide variety of remote-sensing instruments since the early days of space exploration. They have not, as yet, provided complete global cover on an operational basis, nor were they intended to. The presence of trained operators allows experimental systems to be tested from orbit, but the demands on the crews' time from other activities has meant very limited coverage. In many cases the instruments have been directed repeatedly at geologically well-known targets to allow calibration and assessment of the information contained within the data. Except for data collected by Soviet cosmonauts, which are now becoming available, all stem from USA programmes operated by the National Aeronautics and Space Administration (NASA). Archives for many remotely sensed data from staffed NASA missions are held by the US Geological Survey's EROS Data Center and at NASA Goddard Space Flight Center, whose World-wide Web pages appear in Appendix D, along with those of other agencies.

The Mercury and Gemini programmes provided over 1000 normal colour and a few false-colour infrared oblique and near-vertical photographs, taken by hand-held 70 mm cameras. Many of them were directed at geologically interesting areas. Both Gemini and Earth-orbiting Apollo missions were limited to orbits between 35°N and 35°S. As well as 70 mm colour cameras, Apollo 9 (1969) also carried the first orbital multispectral experiment. This consisted of an array of four hand-triggered 70 mm cameras mounted in the port of the command module. Three of the cameras exposed black and white film to give images of the green (0.47–0.62 μm), red (0.59–0.72 μm) and very-near-infrared (0.72–0.90 μm) regions, and the fourth used infrared colour film. Ground resolution proved to be about 100 m. This experiment was directed only at test sites in the southern USA and northern Mexico, and was co-ordinated with aircraft remote sensing flown at the same time as the mission. The experiment was designed as a test of the concept for an unstaffed multispectral scanner, which became the Earth Resources Technology Satellite (July 1972), subsequently renamed Landsat-1 (Section 3.12.2). The success of Landsat-1 prompted NASA to extend its remote-sensing ambitions by a series of experiments aboard Skylab (May 1973 to February 1974), a 100 tonne orbital workshop based on the third stage of a Saturn-V launch vehicle, which orbited between 50°N and 50°S.

3.11.1 Space Shuttle data

NASA's staffed spaceflight programme is centred on the reusable Space Shuttle, the payload capacity and flexibility of which is enormous compared with earlier launch

systems. The second flight in November 1981 (STS-2), with an orbital altitude of 259 km, carried two remote-sensing experiments with geological potential. The Shuttle Multispectral Infrared Radiometer (SMIRR), was a non-imaging spectroradiometer that measured reflectance in 10 wavebands over the range 0.5–2.4 μm , including three with a width of 0.04 μm in the vicinity of the Al–OH feature near 2.2 μm (Fig. 1.9). Its objective was to test the theoretical possibilities for rock and mineral discrimination from orbit (Section 1.3.2).

The second system was the first Shuttle Imaging Radar (SIR-A) experiment. This deployed a 23-cm (L-band) SAR system similar to that on the unstaffed Seasat of 1978 (Section 3.12.1). Unlike Seasat, with its main target being the ocean surface, SIR-A was designed primarily for land applications. To reduce layover problems (Section 3.5) in areas of rugged topography, without producing too much shadowing, SIR-A used a depression angle fixed at 43° to cover a ground swath 50 km wide. Ground resolution was 40 m in both azimuth and range directions. Continuous holographic films representing 8 h of data were correlated optically after the mission to radar images on positive film (Fig. 3.27, see Chapter 7). This record amounts to about 10 million km² of land and ocean between 41°N and 35°S (Appendix D).

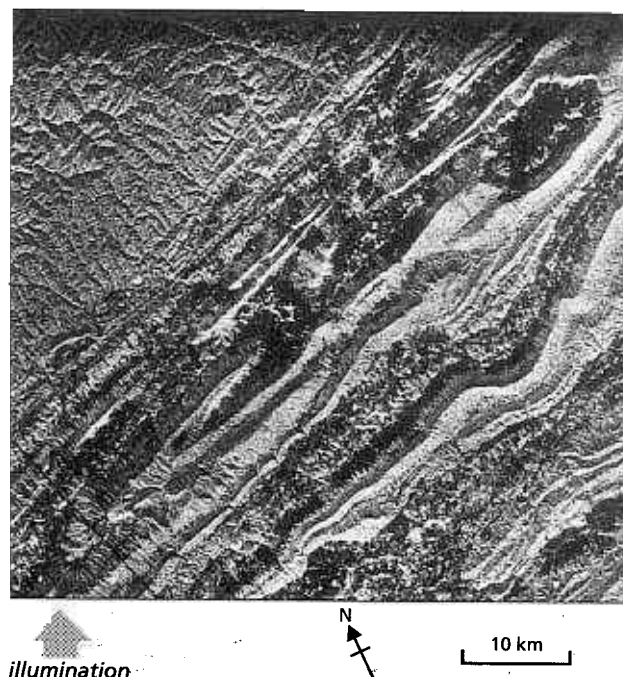


Fig. 3.27 This SIR-A L-band radar image of the Appalachian Plateau and the adjacent fold belt in Kentucky and Virginia shows the advantages over rugged terrain of a low depression angle. There is little distortion and many of the major geological features are displayed well. The same area is shown as a Seasat SAR image in Fig. 3.33. Courtesy of J.P. Ford, Jet Propulsion Laboratory, Pasadena.

In October 1984 the Shuttle carried a system similar to SIR-A, called simply SIR-B, but with a variable depression from 75 to 30° and improved resolution. This resulted in optimum depression angles for different kinds of application, and several images with varying parallax that permit pseudostereoscopic viewing. The pointable antenna meant that test sites could be viewed from a variety of look directions. Data in both optical hologram form and as digital records cover swaths between 57°N and 57°S. Images were expected to be of even better quality than those from SIR-A. Unfortunately, technical problems during the mission thwarted many of the planned acquisitions, and only a few images are of useable quality. The 1994 SIR-C/XSAR experiment, conducted jointly by NASA and the German space agency, involved the use of multispectral radar (L, C and X bands) with various polarizations (Chapter 7), and a polar orbit.

In early 1984, a Shuttle mission carried two European imaging systems that are useful for geological applications. One was a Metric Camera, primarily for cartographic applications, with a ground resolution of 30 m and frame size of 23 cm by 23 cm, which exposed false-colour infrared and panchromatic film. The other was the first test of the German Modular Optical-Electronic Multispectral Scanner (MOMS) which used

a pushbroom system with a ground IFOV of 20 m and two bands (0.58–0.63 and 0.83–0.98 μm). The later MOMS (1996 carried by a Russian staffed satellite) used three pushbroom systems, two pointed fore and aft to give panchromatic stereoscopic capability, and one aimed vertically downwards to gather data with 25 m resolution in four wavebands between 0.45 and 1.05 μm . Neither MOMS provided images of more than a few of the planned scenes.

As well as the SIR-B system, the October 1984 Shuttle mission carried the Large Format Camera (LFC), which produced panchromatic natural colour and false-colour infrared images. The effective resolution of LFC panchromatic photographs is between 8 and 15 m, depending on the altitude of the Shuttle orbits. Each frame covers areas of 250 × 500 to 350 × 700 km, in an archive of 2160 photographs, many of geologically interesting areas in arid and largely unknown terrains. Various overlaps between frames give all LFC images stereoscopic potential. This instrument was never flown again, and NASA was reported to have sold the archive of negatives to a commercial distributor for a peppercorn sum. Figure 3.28 is an example of an LFC photograph for comparison with Landsat images of the same area (Plate 3.3). Several LFC images are used in Chapter 4.

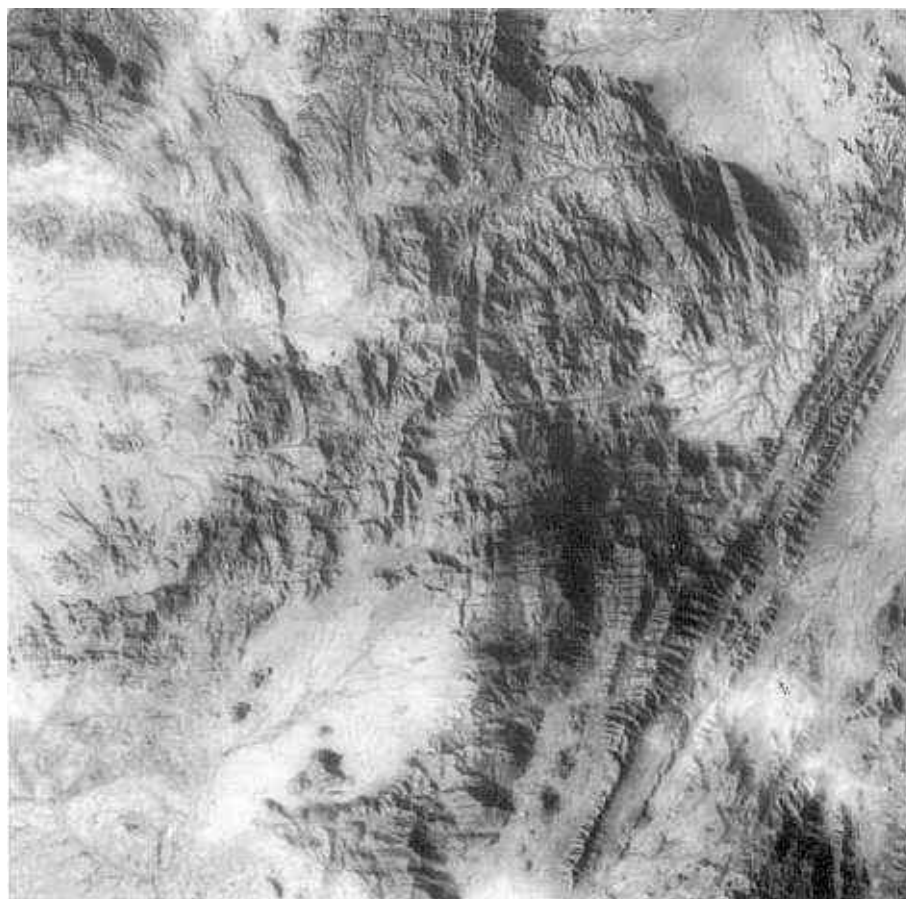


Fig. 3.28 Large Format Camera image of semiarid mountains in Eritrea, north-eastern Africa. The effective resolution is about 10 m. See Plate 3.3 also.

Remote-sensing experiments scheduled for tests aboard the Shuttle until the Challenger disaster included the Thermal Infrared Multispectral Scanner (TIMS), collecting emitted radiation in six bands between 8 and 12 μm (Table 3.2). Results from the aircraft-mounted version of this are discussed in Chapter 6. Another important but unflown candidate was an imaging spectroradiometer similar to AVIRIS (Section 3.9.2, Table 3.3). As in other staffed missions, the astronauts took many photographs from the ports of the Shuttles with a variety of cameras. These are easily obtained from NASA's Goddard Space Flight Center (Appendix D).

3.12 Data from unstaffed spacecraft

There are three advantages of unstaffed orbital platforms.

- 1 By not requiring complex and heavy life-support systems, they are cheap to build and launch, and need not be recovered.
- 2 They can be placed in Sun-synchronous polar orbits that guarantee the possibility of almost complete global coverage (Section 3.10).
- 3 Subject to the deterioration of equipment and exhaustion of propellants required for orbit adjustments, they can operate for very long periods gathering repetitive data. Humans can endure the conditions of orbital travel for only a matter of months.

3.12.1 Meteorological and oceanographic satellites

The most enduring, successful and consistent production of remotely sensed images has been by unstaffed meteorological satellites. To describe and illustrate them all and the variety of devices that they have carried is beyond the scope of this book. Primarily, they monitor global climatic features, with sensors tuned to cloud cover and patterns, atmospheric water content, precipitation and both land- and sea-surface temperatures. Some also carry instruments aimed at oceanographic features, such as the abundance of phytoplankton and particulate matter in sea water. Their role is monitoring the entire planet on a day to day basis. Consequently their design and their orbits ensure coverage of vast areas at coarse resolution, between 800 m and 25 km. For many geological applications the features of interest are too small to be resolved. Their supersynoptic view, however, does serve to gain a gross overview of regional geology. In the same way that high-resolution satellite images help co-ordinate the information drawn from aerial photographs and field work, meteorological images can help draw together the findings from a study of better resolution satellite images.

The first image-gathering unstaffed satellite was the US Television and Infrared Observation Satellite

(TIROS-1), launched on 1 April 1960. It carried two miniature television cameras on an 800-km orbit between about 50°N and 50°S, with a period of about 100 minutes. This began an ongoing series of low-altitude environmental satellites launched from the USA. For administrative reasons a deep moat of confusion surrounds their naming, partly because before launch they had one name and then were rechristened in orbit. Their administering body also changed name. For simplicity they are known as the TIROS/ESSA/NOAA series, which were placed in near-polar orbits to gather data for most of the globe. They were first administered by the US Environmental Science Service (ESSA), which became the National Oceanic and Atmospheric Administration (NOAA).

The first geostationary image-gathering satellite, with an orbit at about 41 000 km, was NASA's Applications Technology Satellite (ATS-1), launched in December 1966. It provided a view of almost an entire hemisphere from a position above the Equator, and was succeeded by a series of similar platforms: Soviet GOMS, the Japanese Himawari (over the western Pacific), the European Space Agency Meteosat (over the eastern Atlantic), and the Indian INSAT (over the Indian Ocean) series. Together these comprise the GOES family, administered by the World Meteorological Organization in Geneva. The bulk of instruments aboard these platforms produce a range of data on a nearly real-time basis, but much is of too coarse a resolution to be useful geologically. The panchromatic visible sensors aboard the geostationary platforms have a resolution of 900 m, however, which does enable some gross geological detail to be seen.

The most geologically useful instrument aboard the current TIROS/ESSA/NOAA series is the Advanced Very High Resolution Radiometer (AVHRR). This is a line-scanning system, with a mirror that sweeps through 56° either side of the nadir of the groundtrack. The system's IFOV together with the orbital altitude (833 km) leads to a ground resolution of about 1.1 km at nadir, degrading outwards because of the large scanning angle. Each scan records data for 2048 pixels, which, together with the off-nadir distortion, means that images cover a swath width greater than 2500 km. All the NOAA series satellites use a near-polar, Sun-synchronous orbit, and cross the same ground area repeatedly at the same local time. Two satellites are always functioning, one passing over at 0730 hours and the other at 1430 hours, on descending orbits (southwards). They cross the night-time side of the Earth 12 h later on ascending orbits. The orbital period of 102 minutes results in 14.1 orbits per day, ensuring complete global coverage every 12 h.

The AVHRR instrument has five channels sampling radiation from the visible red to emitted infrared

Table 3.4 Spectral bands covered by five geologically useful orbiting systems

| Landsat | | AVHRR | | SPOT | | JERS-1 | | ASTER | |
|-----------|-------------------------|-------|-------------------------|------|-------------------------|------------|-------------------------|--------------|-------------------------|
| Band | Range (μm) | Band | Range (μm) | Band | Range (μm) | Band | Range (μm) | Band | Range (μm) |
| MSS 1 (4) | 0.5–0.6 | 1 | 0.58–0.68 | XS 1 | 0.5–0.59 | 1 | 0.52–0.60 | 1 | 0.52–0.60 |
| MSS 2 (5) | 0.6–0.7 | 2 | 0.73–1.10 | XS 2 | 0.61–0.68 | 2 | 0.63–0.69 | 2 | 0.63–0.69 |
| MSS 3 (6) | 0.7–0.8 | 3 | 3.55–3.93 | XS 3 | 0.79–0.89 | 3 | 0.76–0.86 | 3 (+ stereo) | 0.76–0.86 |
| MSS 4 (7) | 0.8–1.1 | 4 | 10.5–11.5 | P | 0.51–0.73 | 4 (stereo) | 0.76–0.86 | 4 | 1.600–1.700 |
| MSS (8) | 10.4–12.6 | 5 | 11.5–12.5 | | | 5 | 1.60–1.71 | 5 | 2.145–2.185 |
| TM 1 | 0.45–0.52 | | | | | 6 | 2.01–2.12 | 6 | 2.185–2.225 |
| TM 2 | 0.52–0.60 | | | | | 7 | 2.13–2.25 | 7 | 1.235–2.285 |
| TM 3 | 0.63–0.69 | | | | | 8 | 2.27–2.40 | 8 | 2.295–2.365 |
| TM 4 | 0.76–0.90 | | | | | | | 9 | 2.360–2.430 |
| TM 5 | 1.55–1.75 | | | | | | | 10 | 8.125–8.475 |
| TM 6 | 10.4–12.5 | | | | | | | 11 | 8.475–8.825 |
| TM 7 | 2.08–2.35 | | | | | | | 12 | 8.925–9.275 |
| | | | | | | | | 13 | 10.25–10.95 |
| | | | | | | | | 14 | 10.95–11.65 |

region (Table 3.4). The detectors output intensity levels in digital form in the range 0–1023 (10-bit precision). Transmission to the ground takes two forms. Data are broadcast continuously at full resolution, when any suitably equipped ground-receiving station can record and display it—High Resolution Picture Transmission (HRPT). Five on-board recorders can each store 10 minutes of full resolution data—Local Area Coverage (LAC)—and a full orbit (110 minutes) of data, resampled to 4 km resolution—Global Area Coverage (GAC). These stored data are transmitted to NOAA ground-receiving stations in the USA, when the satellites are in line of sight with the receiving antennae. Figure 3.29 is an example of a visible-red LAC image from the NOAA-7 AVHRR. One advantage of the daily capture of images from the AVHRR is that eventually a cloud free image or one with minimal cloud cover will be acquired. It therefore becomes possible to mosaic several such images to build up super-synoptic images. This has been done for the whole world, and Plate 3.4 shows a simulated natural-colour infrared image of western Europe and North Africa.

Although suited to investigations of very large geological features, AVHRR imagery is currently under-used by geologists. Estimation of rainfall from cloud density and temperature, and monitoring of snow masses, on a daily basis, provide input to models of stream run-off used by hydrologists. The spectral range means that the Stefan–Boltzmann relationship permits estimation of temperatures from those normally affecting land and sea surfaces, to those associated with active lava flows. Potentially, twice-daily AVHRR data could become the basis for volcano monitoring (Chapter 9).

Between 1964 and 1978 various sensors were tried out on satellites of the Nimbus series, operated by NASA.

Nimbus-7 was the last of the series, launched in October 1978 into a 955-km Sun-synchronous orbit with an inclination of 99.3° and a repeat cycle of 6 days, and was retired in September 1984. Nimbus-7 carried a variety of radiometers ranging from the ultraviolet to microwaves, including the Coastal Zone Color Scanner (CZCS). The CZCS covered a swath width of 1566 km, with a pixel-size of 825 m at nadir. There were six spectral bands, four of which were narrow channels (0.02 μm) in the visible region for mapping algal chlorophyll and other suspended material in the oceans, a VNIR channel for mapping surface vegetation and a thermal infrared channel aimed at ocean temperature variations. The CZCS data had great potential for regional geological overview, but have rarely been used because of the patchy nature of its cover. In 1997 NASA replaced the CZCS with a similar, but finer resolution system known as the Sea Wide-field Sensor (Sea-WIFS) (Appendix D).

Seasat was launched into a 790-km, 108° inclination polar orbit in June 1978. It carried a variety of radiometers and a radar altimeter (Chapter 8), but more significantly it had the first SAR system to be used in orbit. This was an L-band (23.5 cm) system with a depression angle of 70° giving a swath width of 100 km and a resolution of about 25 m. A steep depression angle results in topographic layover (Section 3.5) and other distortions in rugged terrains (Fig. 3.30), but is especially suitable for imaging the ocean surface to assess its roughness. In fact, as its name suggests, Seasat was designed for oceanographic applications, but the radar also proved useful over land. The area of SAR coverage for Seasat was limited to North America and the northern Atlantic region by the available ground stations. Unfortunately a power failure terminated Seasat operations after only



Fig. 3.29 The gross geomorphology of an area of about 4 million km², covering parts of Afghanistan, Pakistan, part of the former USSR and Tibet is shown on this Channel 1 (red) image from the AVHRR aboard NOAA-7. The High Himalaya and Tibetan Plateau in the east are largely snow- or cloud-covered. The Himalayan foothills are densely forested and appear dark. The fertile plains of the Indus and its tributaries are dark, in contrast to the light coloured desert areas. The geological advantages of a supersynoptic view are amply illustrated by the interconnected complexities of the Sulaiman and Salt Ranges of north-west Pakistan. They are truncated in the far west by a huge strike-slip fault, along which the Indian Subcontinent has slipped during its collision with Asia.

105 days of operation. Nevertheless it provided sufficient evidence of the advantages of all-weather, day or night orbital SAR imaging to increase the impetus for more geologically orientated systems, such as the SIR series.



illumination  **10 km**

Fig. 3.30 This extract from a Seasat L-band radar image is of the same area in north-east USA as Fig. 3.27. The radar look direction is at 45° to that for the SIR-A image in Fig. 3.27. This, combined with the steeper depression angle for the Seasat SAR compared with that for SIR-A, makes the image virtually unintelligible. The look-direction is parallel to the main mountain ridges and valleys in the Appalachian fold belt and so they are suppressed. The rugged topography has resulted in the image being plagued by foreshortening and layover effects, which further disrupt its structure. However, in the Appalachian Plateau (northern part of the image) there are advantages. The dominant southwards drainage is enhanced, allowing discordant linear features, probably major faults, to be picked out.

3.12.2 The Landsat series

Experiments from staffed satellites and the success of the semipermanent meteorological satellite systems encouraged NASA to initiate a series of high-resolution orbital systems aimed at the biological and physical resources of the Earth. The first Earth Resources Technology Satellite (ERTS-1) was launched by NASA on 23 July 1972, using a Nimbus platform (Fig. 3.31a). It was subsequently renamed Landsat-1, and eventually the administration of the Landsat series passed first to NOAA and then to a commercial company EOSAT (now Earth Imaging) (Appendix D). The 918 km Landsat orbits passed within 9° of the poles, and circled the Earth once every 103 minutes. Descending orbits crossed the Equator at an angle of 9° at around 0930 hours local time. So, the Landsat series was Sun-synchronous, to give illumination of the surface by the mid-morning Sun. This avoids early morning cloud at mid-latitudes, but

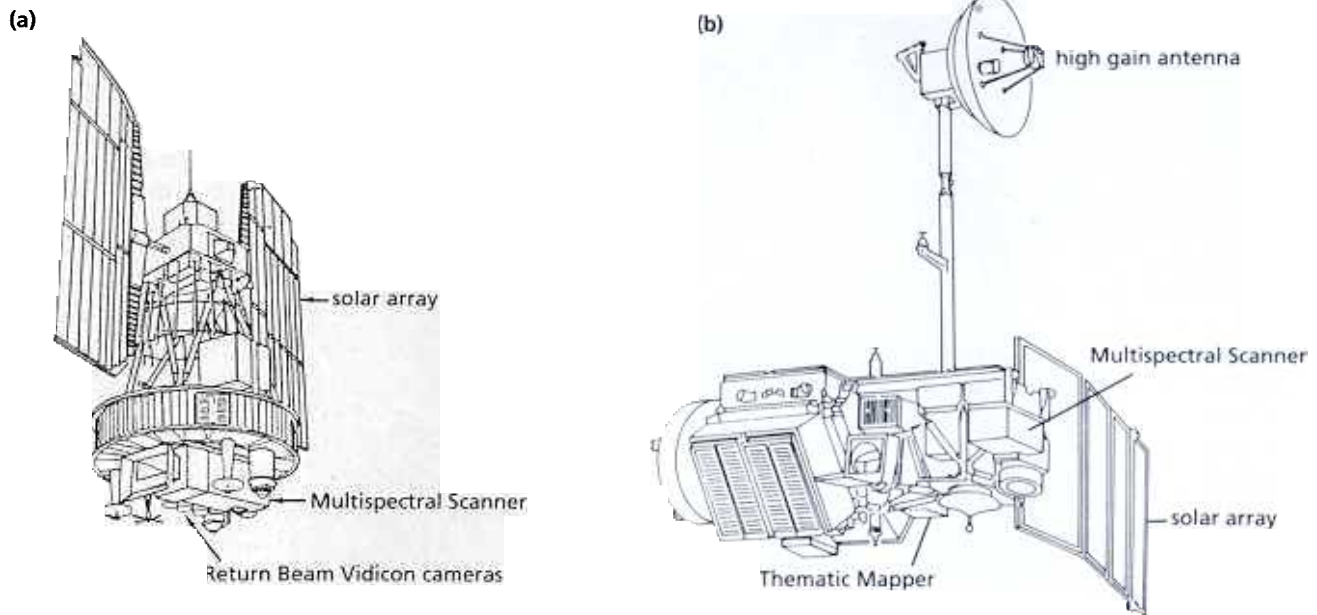


Fig. 3.31 The first three Landsats occupied spare Nimbus platforms (a), whereas that for Landsat-4 and -5 is a modified TIROS platform (b).

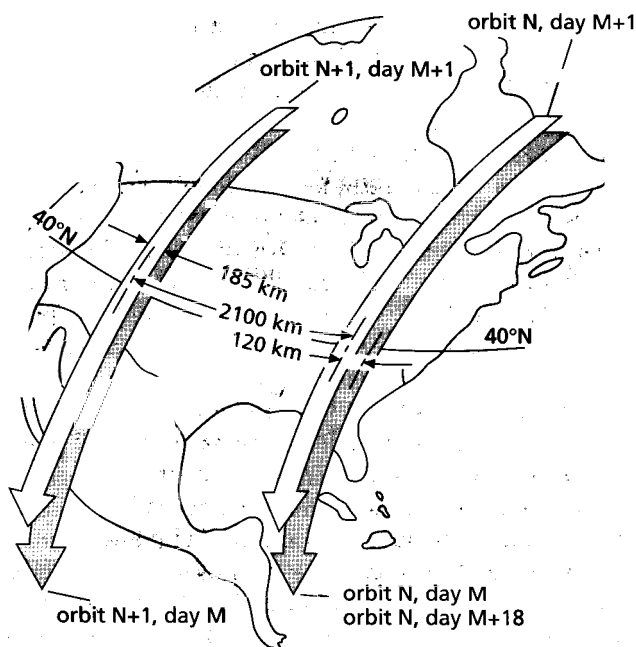


Fig. 3.32 The diagram shows successive Landsat orbits on two days. On day M, orbit N + 1 is shifted 2100 km west of orbit N at this latitude. Orbit N on day M + 1 is only 120 km west of the same orbit of the previous day, so the images taken on both days overlap. The higher the latitude the greater the overlap between adjacent scenes. The sequence repeated itself every 18 days for Landsat-1, -2 and -3. This revisit frequency was reduced to 16 days for Landsat-4 and -5. Very much the same principle applies to all Sun-synchronous polar-orbiting systems. In the case of the NOAA AVHRR, the revisit frequency is every 12 h to give day and night coverage.

also highlights topography because of the low Sun angle. Successive orbits were 2760 km west from their predecessor (Fig. 3.32), so Landsat-1 crossed the same point on the ground every 18 days. Every part of the globe, except for the near-polar regions, which are never crossed, can therefore be imaged up to 20 times per year. Unlike the AVHRR, Landsat imaging systems cover a narrow (185 km) swath, so there are gaps in the ground coverage in successive orbits (Fig. 3.32). Over an 18-day period (251 orbits), however, Landsat-1 to -3 gave a complete set of ground tracks, spaced at 159 km apart at the Equator and giving a 14% sideways overlap, after which time the cycle repeated. The sideways overlap increased towards the poles, and reached 50% at 54° latitude, so that every point at this latitude was covered twice as frequently as most points on the Equator. The ground track of a Landsat-1 to -3 satellite over a single day is illustrated in Fig. 3.26, and the relationship between adjacent orbits on successive days is shown in Fig. 3.32. This repeated pattern of overpasses, common to all Sun-synchronous satellites, is the basis for the Landsat world-wide reference system for individual scenes, the principle of which is followed by other operational systems.

In 1982 NOAA adopted a new orbital plan for Landsat-4 and its successors. This involved a new, heavier platform (Fig. 3.31b) and a lower orbit (705 km), resulting in a shorter orbital period (99 minutes), a shorter repeat cycle of 16 days and a greater number of orbits to give global coverage. This was to give greater orbital stability and to help improve the resolution of the on-board sensors.

Landsat-1 and -2 were equipped with two imaging systems. One was a three-channel Return-beam Vidicon (RBV) system, the other a four-channel Multispectral Scanner (MSS), both aimed at visible and VNIR bands containing features related to chlorophyll (Table 3.4). The original Landsat concept was orientated towards monitoring of vegetation rather than geology. The one concession to geologists was the low Sun-angle illumination, to highlight appropriately orientated topographic features. There can be little doubt that Landsat was funded in order to maintain a permanent high-precision watch on world agriculture, particularly that of the USA's main customer for wheat, the former Soviet Union. Nevertheless, the quality of the early Landsat data found exciting applications in geology.

The RBVs aboard Landsat-1 and -2 provided little data. However, they left a legacy in the naming of the four MSS bands on Landsat-1 to -3, which were designated bands 4, 5, 6 and 7 because they followed the RBV bands in sequence. The Landsat-1 to -3 MSS system had a scanning angle of about 12° using an oscillating mirror (Fig. 3.8). This collected radiation for banks of six sensors for each of the four wavebands, which were directed through a dispersing element to the detectors by optic fibres. The dimensions of the fibre ends determined the IFOV of the system, which was $109 \mu\text{rad}$, resulting in a ground resolution of $79 \times 79 \text{ m}$. The sampling rate, however, was such that each pixel overlapped those adjacent by about 11.5 m, resulting in an actual pixel size of 79 m depth and 56 m width. With the lower orbit of Landsat-4 and -5 the MSS scan angle was increased to maintain the swath width at 185 km, and the optics were adjusted to preserve the approximately 80 m IFOV (actually 82 m). The spectral bands were identical, although renumbered as bands 1–4 as mentioned above, so that old and new MSS data are very closely comparable to maintain data continuity.

The analogue data from each MSS detector are electronically chopped and digitized to a 0–63 range of DN (6-bit precision). On Landsat-1 to -3 the data were either relayed directly to a ground-receiving station or recorded until the platform was above a ground-receiving station. On Landsat-4 and -5 there are no tape recorders and data are telemetered to the ground either directly or via a series of tracking and data relay satellites (TDRS). The data are rescaled on the ground to 0–127 for bands 4, 5 and 6 (now 1, 2 and 3) and 0–63 for band 7 (now 4), and output either to film using a digital film writer, or to digital storage. In both cases the data use the range 0–255 (8-bit precision), but of course only 128 different real levels are possible in the data. The detectors are calibrated by lamps built into the system in such a way that the darkest expected surface—deep clear water or black rock—and the lightest—cloud or snow—both give responses within the range. Because

cloud or snow is so much brighter in the four bands than any surface vegetation, soil or rock, this means that cloud-free images generally have their data compressed into a relatively small range of low digital numbers and require contrast enhancement (Chapter 5).

The Landsat-1 to -3 MSS collected data for 3240 pixels 56 m wide during every sweep of the mirror, each 474 m deep—79 m for each of the six detectors. The ground swath imaged on each overpass therefore was about 185 km wide. The continuous stream of data was sampled about every 25 s, giving a total of 2340 lines, so that each sample represents an area $185 \times 185 \text{ km}$ on the ground. This is termed a scene. Each scene acquired by Landsat does not represent a square area on the ground, but a parallelogram (Fig. 3.11h), because of the effect of the Earth rotating in the time taken to gather the data. Neither are the scenes orientated precisely relative to true north. This is because the orbits are inclined to latitude; in fact they are slightly curved (Fig. 3.26). Apart from data from Landsat-1 and -2, scenes are corrected for Earth rotation effects, although the user must register them to local map co-ordinates (Appendix B). An extract from a Landsat MSS image is shown in Plate 3.3(a).

The other addition to Landsat-3 was a line-scan system for emitted infrared. This thermal data had a resolution of 120 m, with a similar structure to the MSS data. However, the thermal channel failed shortly after launch.

The gradual deterioration in the performance of all systems aboard Landsat-3 forced an early launch of its successor, Landsat-4 in September 1982. As well as a platform design based on that of the TIROS series (Fig. 3.31b) and the replacement of onboard recording capability with direct telemetry of data by multipurpose data-relay satellites, Landsat-4 carried an entirely new instrument—the Thematic Mapper (TM)—as well as a slightly modified MSS. The TM is a line-scan system, but incorporates several innovations compared with MSS. Instead of using fibre optics to carry radiation to the detectors, they are placed directly at the focal planes of the optical system. One of these focal planes carries four banks of 16 detectors for four wavebands in the visible to VNIR range. Another is refrigerated by a radiative cooler, and carries two banks of 16 detectors for two wavebands in the SWIR region, and one bank of four detectors for emitted infrared. Table 3.4 shows the specifications for TM wavebands, where there is an apparent anomaly in the numbering of the bands—band 7 is in the SWIR, and band 6 contains thermal data. This arose from the late addition of the SWIR band at the request of a vociferous geological lobby. The arrangement allows the mirror to scan a ground strip with approximately the same width as that scanned by the MSS, but radiation is collected by 16 instead of six detectors for the visible and near-infrared channels.

The ground IFOV in these channels is therefore 30 m, an improvement of about 2.6 times compared with that of the MSS. Resolution for the thermal channel is 120 m. To match the IFOV to the platform's orbital velocity, the mirror gathers data in both forward and backward sweeps. This means that the actual scan lines are not exactly parallel on the ground because of orbital motion. Correction for this distortion results in an actual pixel size of 28.5 m.

Thematic Mapper scenes are almost the same size as those from the accompanying MSS, but because of the improved resolution and larger number of channels they comprise 8.5 times as many digital numbers. Landsat-4 ceased to operate early in 1983 because of problems with power supply and serviceability of its relay system. Fortunately it was possible to replace it with the already prepared, but slightly modified Landsat-5 system, in March 1984. Plate 3.4 compares a Landsat TM image with one from the MSS system. Landsat-5 continues to function, with Landsat-4 as a back-up, but both are long past their planned lifetime. Their replacement, Landsat-6, ignominiously fell into the sea after lift off. After much wrangling, Landsat-7 with an additional 15 m panchromatic band, was launched successfully in 1999. At the time of writing NASA was about to launch an experimental, low-cost platform to test components for future members of the Landsat series.

Because of the overlap between the image swaths for adjacent Landsat paths, which increases with increasing latitude, it is possible to view parts of the images stereoscopically (Appendix A). A system designed to produce stereoscopic pairs of images by pointing the detecting system off-nadir, thereby introducing parallax differences between images, is the French *Système Probatoire de l'Observation de la Terre* (SPOT).

3.12.3 SPOT data

The French Centre National d'Études Spatiales (CNES) launched SPOT-1 into an 830-km Sun-synchronous polar orbit (98.7° inclination) in February 1986. The SPOT system is operated by SPOT-Image. The platform and its successor SPOT-2 cross a point on the Equator on descending orbits at 1030 hours solar time every 26 days. The SPOT system uses two pushbroom devices (Section 3.4) (Haute Resolution Visible—HRV) comprising arrays of 6000 CCDs (Figs 3.13 and 3.33a). These may operate in either a panchromatic mode with a ground IFOV of 10 m, or multispectrally to gather green, red and VNIR reflectance (Table 3.4) with an IFOV of 20 m. The imaged swath, when directly below the platform, is 60 km wide. In the panchromatic mode scenes comprise an array 6000 pixels square, and in the multispectral mode each scene is 3000 pixels square.

The SPOT system marked two significant advances in remote sensing; it was the first unstaffed satellite to use a pushbroom system, and was the first remote sensing satellite to offer stereoscopic cover. By pointing both HRV sensors downwards it is possible to image a strip 117 km wide, with a 3-km overlap between instruments (Fig. 3.33b). Alternatively, by means of steerable mirrors that direct radiant energy on to the CCD arrays and point up to 27° either side of vertical, one or both HRV instruments can image a swath centred up to 475 km on either side of the ground track. This ability means that a stereoscopic pair of images can be acquired by imaging the same area from different directions, during different orbits (Fig. 3.33c). A stereoscopic SPOT image is used to illustrate part of Section 4.4.1. The same area can also be revisited up to 11 times during the 26-day period, depending on the latitude (Fig. 3.33d). As well as providing routine high-resolution images, SPOT is also capable of responding to short-lived events, such as floods and volcanic eruptions.

The SPOT platforms carry a tape recorder and transmit images recorded from around the world, as well as images of Europe in real time, to ground stations at Kiruna (Sweden) and Toulouse (France). In addition, many of the Landsat ground stations are also licensed to record and distribute SPOT data. The next two in the series, SPOT-2 and -3 were the same as the first two systems. SPOT-4 carries an additional spectral band at 1.58–1.75 μm and also a 1–4 km pixel vegetation-sensing system, using five spectral bands (blue, green, red, VNIR and SWIR), with a swath 2200 km wide (extending 50° either side of nadir). SPOT-5 carries improved sensors capable of 10 and 5 m resolution for multispectral and panchromatic modes, respectively.

3.12.4 JERS-1 data

In February 1992 the Japanese National Space Development Agency (NASDA) launched what was then the most technically sophisticated, multipurpose remote sensing system in the civilian domain. It was called simply the Japanese Earth Resources Satellite-1 (JERS-1), and carried two imaging systems, an Optical Sensor (OPS) and a synthetic aperture radar system (SAR). The JERS-1 orbited at an altitude of 568 km with a 97.7° inclination, which allows a revisit period of 44 days.

The OPS covered the reflected region of the spectrum, being intended for detailed discrimination of vegetation, soils and rocks. It used two radiometers, one for the visible and near-infrared regions, the other for the SWIR. Both pointed vertically downwards, but that carrying visible and near-infrared sensors could also point 15.3° ahead of the nadir to obtain stereoscopic images. The seven spectral bands capturing data from the nadir

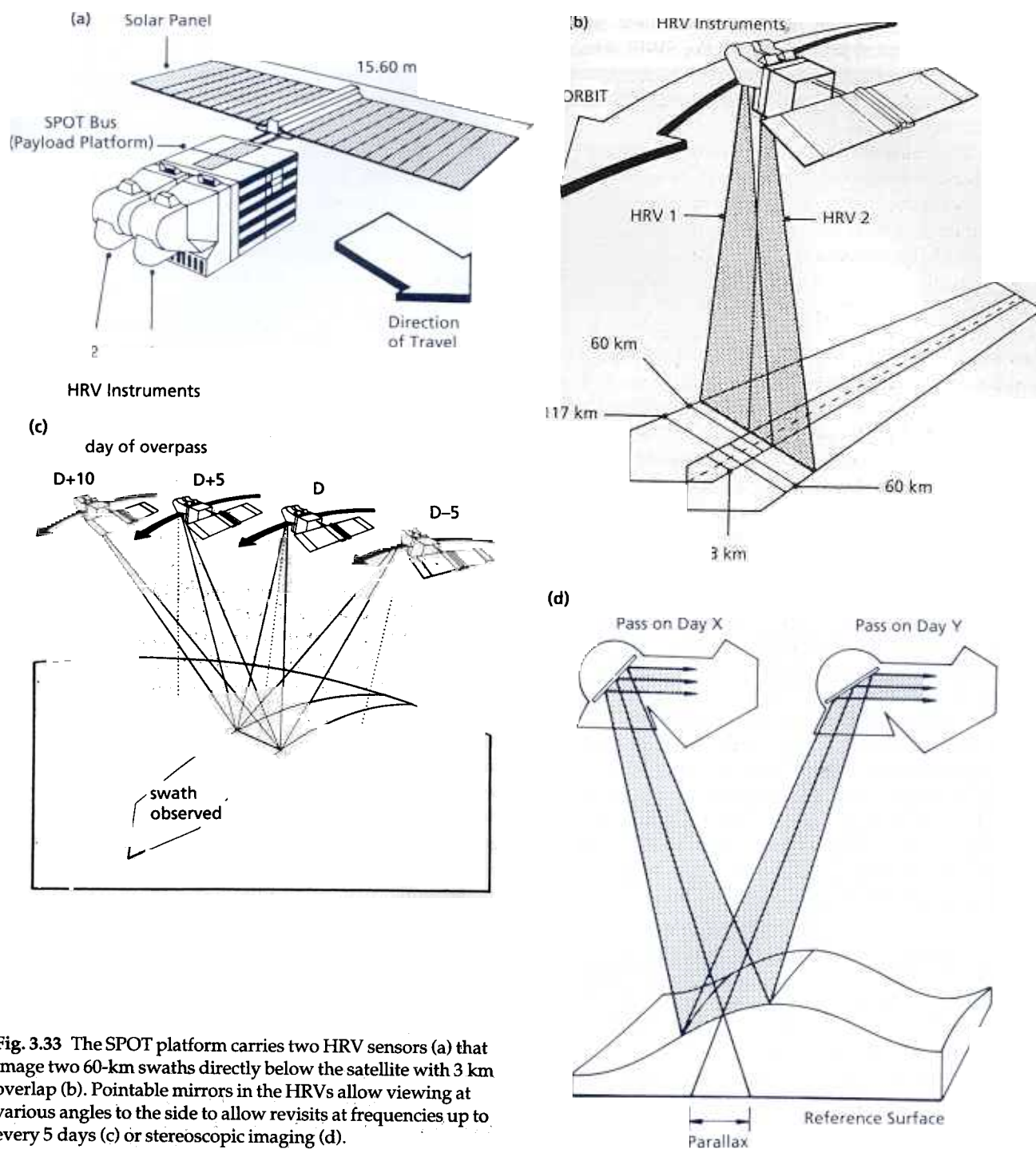


Fig. 3.33 The SPOT platform carries two HRV sensors (a) that image two 60-km swaths directly below the satellite with 3 km overlap (b). Pointable mirrors in the HRVs allow viewing at various angles to the side to allow revisits at frequencies up to every 5 days (c) or stereoscopic imaging (d).

(Table 3.4) each had a dedicated 4096 element CCD array capable of measurements with 6-bit (64 DN) precision. This configuration, together with an IFOV of $32.2 \mu\text{rad}$, produced images 75 km across with a pixel size of $18.3 \times 24.2 \text{ m}$. The forward-looking radiometer used only a single VNIR band to correspond with one from the nadir-pointing mode for stereoscopic imagery.

The three bands in the $2.0\text{--}2.4 \mu\text{m}$ region were aimed specifically at Al–OH, Mg–OH and C–O bond transitions (Section 1.3.2) and should have enabled more confident separation of clays, hydroxylated ferromagnesian minerals and carbonates than had been possible before. Technical considerations limited the dynamic range of the data to 64 DN compared with the 256 DN of Landsat

TM data, which trimmed this optimistic view somewhat. Sadly, technical problems with the SWIR sensors introduced seriously disturbing noise, which required time-consuming cosmetic processing before JERS-1 data were useable.

The JERS-1 SAR also had potential for geologists. Its 23 cm L-band system had a depression angle of 55° aimed at suppressing the layover and shadowing in high-relief areas that plague SAR images with larger and smaller depression angles, respectively (Fig. 3.22).

3.12.5 Other systems of interest

The Heat Capacity Mapping Mission (HCMM) experiment launched by NASA in April 1978 had a near-polar orbit at 620 km (97.6° inclination). It ceased operation in September 1980. The HCMM orbits were Sun-synchronous with times of equatorial passage at 1400 and 0300 hours local time, chosen to match those of the average maximum and minimum daily soil temperatures. Potentially, it could estimate variations in surface thermal inertia, when the day and night-time data were registered (Section 1.3.2). It carried two sensors, one covering visible and near-infrared (0.5–1.1 μm) regions, the other the emitted infrared (10.5–12.5 μm) region. The ground resolution of the thermal scanner was about 600 m, reduced to 480 m after onboard processing. The imaged ground swath was about 716 km wide. Because day and night orbits were inclined to one another, each scene of registered thermal data was lozenge-shaped. Without an onboard recording system, data are available only for areas within range of suitably equipped ground-receiving stations. The potential for extracting important information relating to rock types and to soil moisture is sufficiently great, however, that HCMM results are discussed at some length in Chapter 6.

The first Indian Remote Sensing Satellite (IRS-1 A) was launched into a 904-km Sun-synchronous polar orbit in March 1988 by a Russian rocket. Its orbital repeat was 22 days, with an equatorial overpass time of 1000 hours. The satellite carried three pushbroom Linear Imaging Self Scanning (LISS) cameras, each recording the same four visible and near-infrared spectral bands (0.4–0.5, 0.5–0.6, 0.6–0.7 and 0.7–0.9 μm). LISS-1 covers a 148-km swath with a ground IFOV of 73 m. LISS-2 A and LISS-2B are identical and image adjacent swaths on either side of the ground track at 36.5 m pixel size. Most data are transmitted to the Indian ground station in Hyderabad (limiting the coverage to India and its vicinity), but there are plans to receive some data at other receiving stations. The latest, IRS1-C, also carries a 5-m resolution panchromatic, pushbroom sensor, providing useful, medium-cost image data for several geological applications.

In 1988 images from a part of the Soviet space programme, which had hitherto been shrouded in secrecy, came on to the market. They stem from a satellite series known as Resurs, the first of which became operational in 1980 and which are in 250 km polar orbit. The Resurs images can be obtained only in photographic form, although they are scanned to produce digital images, and they are captured by four sensor systems. Three of these are line-scanners operating in the visible and near-infrared (Kate-140, Kate-200 and MKF-6). The fourth system is high-resolution (5 m) photography, using colour infrared film, ejected from the satellite and recovered on the ground (KFS-1000). Some of the data have stereoscopic potential. However, permission from the authorities of an overflown country is needed to gain access to the data.

In 1991 the European Space Agency (ESA) launched ERS-1 in a blaze of publicity. Much of this was aimed at the geological community because ERS-1 carries a 5.3-cm C-band SAR imaging system, as well as a number of other instruments for marine applications. ERS-1 flies in polar orbit at 777 km altitude, but is not Sun-synchronous. The SAR imagery from ERS-1 has an 80-km swath width and is rated at a 30-m resolution for range and azimuth resolution. Its depression angle is the same as that for Seasat (67°), as it is largely orientated to monitoring the state of the ocean surface. Such a steep depression inevitably leads to layover in even moderately rugged topography. Although ERS-1 SAR imagery is of extremely good quality, the layover problem serves to highlight the deficiencies inherent in its design (Fig. 3.34a). It seems to offer little advantage in geological mapping in all but the most muted terrain because of this. Figure 3.34(b) shows a JERS-1 SAR image of the same area, where the terrestrially more appropriate depression angle produces geologically better information. However, ERS-1 and -2 are highly productive, and special processing that takes advantage of some of the properties of radar signals allows very precise measurement of terrain elevation. As well as being useful in cartographic applications, by providing digital elevation models (Chapter 8), the precision is so high that comparing elevations from different dates gives accurate measurements of small changes in surface elevation. That is discussed in Chapters 8 and 9.

Canada and the USA launched Radarsat in 1996, as a multipurpose imaging radar system. It deploys a 5-cm SAR system that operates in several modes that cover swathes from 45 to 500 km across, with resolutions from 10 to 100 m. It also uses variable depression angles to optimize its performance over different terrains. Radarsat's pointability allows monitoring of short-lived phenomena, too. Detection and monitoring of Arctic sea-ice was its primary goal, but its flexibility gives it a multipurpose role.

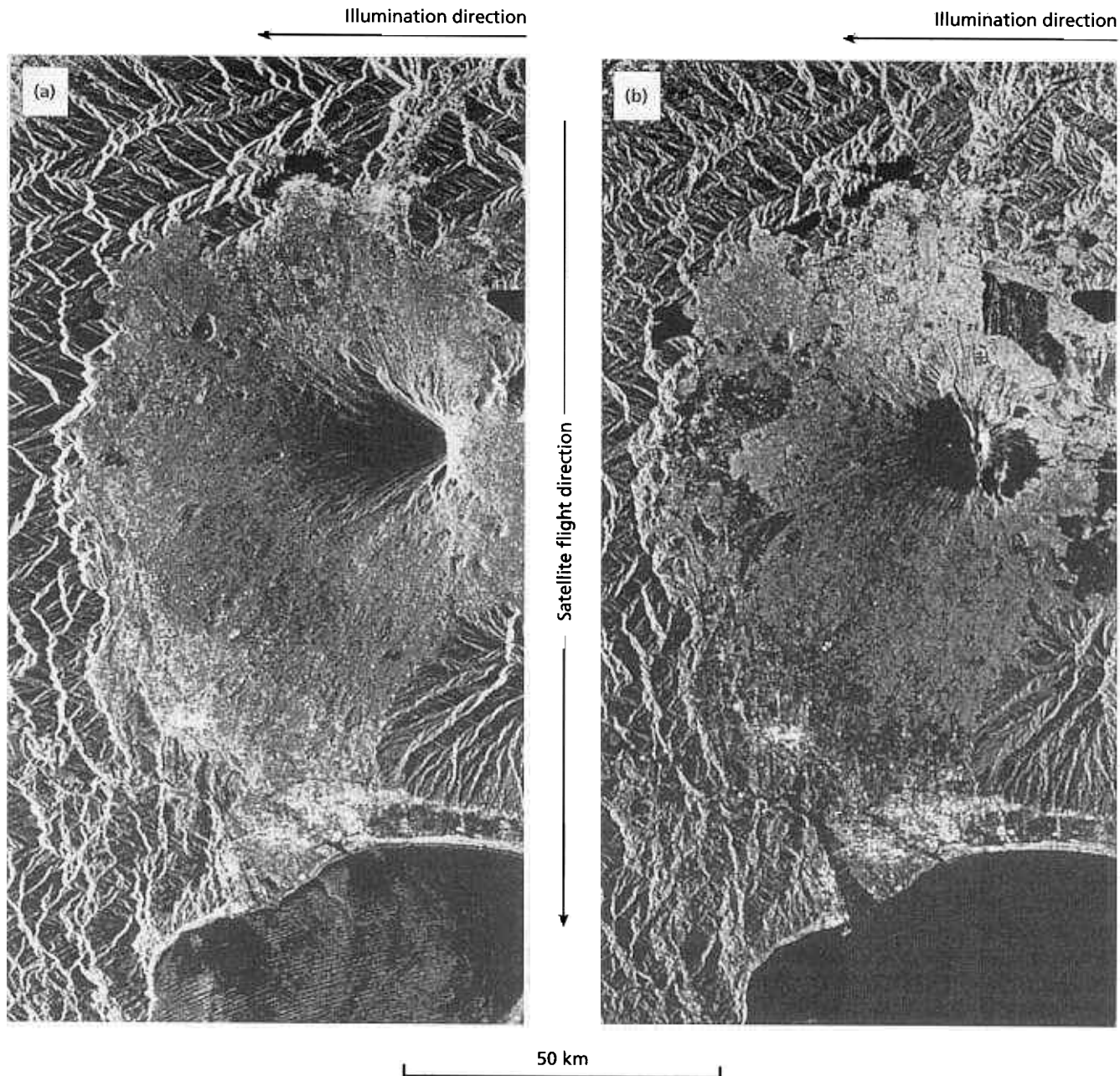


Fig. 3.34 ERS-1 (a) and JERS-1 (b) SAR images of part of Japan, showing the volcano Mount Fuji. The fact that Mount Fuji is a nearly perfect cone with a circular summit crater serves to demonstrate the inappropriate depression angle of ERS-1 SAR

by its apparently lying on its side. Many other rugged topographic features are also completely distorted by extreme layover. The JERS-1 image preserves the shape of the volcano, but still contains layover.

3.13 Future prospects

For the period up to 2005 there are plans for enough civilian remote-sensing satellites to fill a separate chapter. Plans, however, only materialize given money and successful launches. Owing to technical difficulties, even funded ventures do not leave the ground on time. The first of the Earth Observation System platforms (below), which is to carry the ASTER instrument, eventually

reached orbit after an 18-month delay (December 1999). Some do not work properly, leading to near-suicidal disappointment among scheduled investigators, who may have planned for years. Limiting the scope of this section is consequently not so difficult. I discuss only those future systems that have a geological impact, and which seem likely to fly.

The early years of the 21st century seem likely not to see great innovation in land-orientated remote sensing, except perhaps for the wider availability of hyperspectral

data. The Lewis satellite, which was to carry a 384-band imaging spectrometer, failed to achieve orbit in 1998. Perhaps that is just as well. The vast bulk of geologists are not 'up to speed' with Landsat, let alone with hyperspectral capabilities of imaging spectrometers. The same cannot be said for atmospheric and oceanographic studies, which are far more advanced. The growing problems of global warming, pollution and environmental change have forced a change of focus from simple meteorological data to the application of arcane physical laws in the quest for means of understanding what the Earth is currently going through. For most users of land-orientated remotely sensed data, the most pressing need is for assured continuity of tried and tested systems. The Landsat series is guaranteed up to the launch of the seventh platform in the series, with a few modifications, and SPOT and IRS are similarly ensured.

The 1990s witnessed great activity, with new ventures into multifrequency, multipolarization and selectable depression angles for SAR imagery, imaging spectrometry and multispectral thermal infrared systems, all tested using airborne platforms. Apart from the possible development of active systems aimed at exciting mineral fluorescence and improvements in gamma-ray spectrometry, there seems little likelihood of entirely new developments in the foreseeable future. Instead, much aspiring research depends on the carrying of experimental systems into orbit for further testing before truly operational satellites are launched. These ambitions centre on the joint USA/European/Japanese Earth Observing System (EOS) project, which plans up to three unstaffed polar-orbiting platforms starting in 1999.

The bulk of plans for the EOS are not high-resolution land-orientated systems, but three are of immediate interest to geologists, although not all are equally assured. The first EOS platform (EOS AM-1, renamed Terra after a public competition in the USA for an imaginative name) successfully entered orbit on 18 December 1999, where it carries a development from the principle of the Geoscan and JERS-1 systems; that of using a few narrow wavebands strategically placed around mineralogically useful spectral features. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) incorporates three spectrometers; one for the visible and VNIR with three bands (15 m resolution), one for the SWIR region with six bands (30 m resolution), and one aimed at mineralogical features in the thermally emitted part of the spectrum with five bands (90 m resolution). The ASTER also uses a forward-pointing VNIR sensor to provide stereoscopic imagery. Details of the band positions and widths are in Table 3.4. At the time of writing ASTER had produced no scientific data, but was functioning properly. The goal of the mission is to generate cloud-free cover for the entire continental surface during

its 6-year planned lifetime; it could supplant Landsat for many geological applications.

The AVIRIS imaging spectrometer has had such success that a similar instrument—the High Resolution Imaging Spectrometer (HIRIS)—seems an ideal candidate for EOS, particularly with the unseemly demise of Lewis. Technical constraints in planning, mainly to do with weight and power requirements, however, have resulted in its deferment to later EOS platforms, which are not yet assured. The HIRIS will incorporate 192 bands (Table 3.3), sufficient for adequate discrimination of many important spectral features. The images will be 800 pixels wide at a resolution of 30 m, giving a swath 24 km wide. Similar technology is incorporated in the Moderate Resolution Imaging Spectrometer (MODIS) but this has a 500-m ground IFOV that makes it of marginal interest to geologists. The MODIS is aboard EOS AM-1 (Terra). In April 2000, NASA launched a low-cost satellite (EO-1) to test follow-on instruments for the Landsat series, but it also carries a narrow-swath (7.5 km) hyperspectral system called Hyperion.

One of the experimental victims of the aftermath of the Challenger disaster was lengthy delay in the testing of a follow-on to SIR-B, the SIR-C/XSAR system with 3.0, 5.3 and 24 cm wavelengths, which could operate in four different polarization modes and with variable depression angles. Experiments with multispectral, multipolarization SAR systems, such as SIR-C/XSAR, have had mixed success in geological applications. Such a system was proposed initially for EOS, but is in limbo at the time of writing. There are considerable problems with SAR systems in orbit, not the least of which are weight, high power consumption and high demands on data transmission. NASA launched a Space Shuttle Mission dedicated to measuring topographic elevation over most of the land surface using radar interferometry (Chapter 7) (Shuttle Radar Topography Mission—SRTM) in February 2000. As discussed in Chapter 8, digital elevation models (DEMs) are sources of an immense amount of information relating to terrain. That from the SRTM will have a spatial resolution of about 30 m, with an elevation precision of 10 m, and promises to add a new direction in many geological studies.

Aerial photographs continue to be the workhorses for geological applications, because of their high resolution, as you will see in Chapter 4. However, costs are rising and storage of film has always posed problems. Similar resolutions are now achieved by entirely digital cameras, and the data can be stored compactly on CD-ROMs as well as being designed for computer analysis. The same technology potentially has a large market for illustrating news items, and several commercial outfits plan to launch 'mediasats' with resolutions of the order of 1 m. Most, such as Earthwatch, Eyeglass and Space Imaging, will capture stereoscopic images to serve increasingly

sophisticated cartographic and planning requirements. Clearly they will find a ready use among geologists—if they are cost-effective—particularly for areas where it is difficult to arrange aerial surveys.

Geologists seem to have rosy prospects in remote sensing for the next decade. This period is likely to be one of consolidation rather than innovation, giving the majority of geologists the time to get to grips with what has been happening over the last three decades in geological remote sensing research, to apply the new data to exciting new geological problems instead of repeatedly poring over tiny test areas, and to catch up with their colleagues in other fields. The first step along this road is to review the basic methods of image interpretation using simple photographic data from both aircraft and satellites in Chapter 4.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file Menu.htm in the Imint folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

On the Menu page, select the link Miscellaneous: Image data, which will take you to a page that gives examples of many kinds of image.

Further reading

- Allison, L.J. & Schnapf, A. (eds) (1983) Meteorological satellites. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 651–679. American Society of Photogrammetry, Falls Church, Virginia.
- Chavez, P.S. & Bowell, J.A. (1988) Comparison of the spectral information content of Landsat Thematic Mapper and SPOT for three different sites in the Phoenix, Arizona region. *Photogrammetric Engineering and Remote Sensing* **54**, 1699–1708.
- Chiu, H.Y. & Collins, W. (1978) A spectroradiometer for airborne remote sensing. *Photogrammetric Engineering and Remote Sensing* **44**, 507–517.
- CNES (1982) *SPOT Satellite-Based Remote Sensing System*. CNES, Toulouse.
- Doyle, F.J. (1985) The Large Format Camera on Shuttle mission 41-G. *Photogrammetric Engineering and Remote Sensing* **51**, 200.
- Elachi, C. (ed.) (1983) Microwave and infrared satellite remote sensors. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 571–650. American Society of Photogrammetry, Falls Church, Virginia.
- Freden, S.C. & Gordon, F. (1983) Landsat satellites. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 517–570. American Society of Photogrammetry, Falls Church, Virginia.
- Goetz, A.F.H. & Herring, M. (1989) The high resolution imaging spectrometer (HIRIS) for EOS. *IEEE Transactions on Geoscience Remote Sensing* **GE-27**, 136–144.
- Goetz, A.F.H., Vane, G., Solomon, J.E. & Rock, B.N. (1985) Imaging spectrometry for earth remote sensing. *Science* **228**, 1147–1153.
- Hiller, K. (1984) MOMS-01 Experimental Missions on Space Shuttle Flights STS-7 June '83, STS-11 (41-B) February '84. *Data Catalogue*. DFVLR, Overpaffenhofen.
- Jensen, H., Graham, L.C., Porcello, L.J. & Leith, E.N. (1977) Side-looking airborne radar. *Scientific America* **237**, 84–95.
- Kahle, A.B., Shumate, M.S. & Nash, D.B. (1984) Active airborne infrared laser system for identification of surface rock and minerals. *Geophysical Research Letters* **11**, 1149–1152.
- Lowe, D. (1980) Acquisition of remotely sensed data. In: *Remote Sensing in Geology* (eds B.S. Siegal & A.R. Gillespie), pp. 48–90. Wiley, New York.
- Lowman, P.D. (1980) The evolution of geological space photography. In: *Remote Sensing in Geology* (eds B.S. Siegal & A.R. Gillespie), pp. 91–115. Wiley, New York.
- Moore, R.K., ed. (1983) Imaging radar systems. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 429–474. American Society of Photogrammetry, Falls Church, Virginia.
- NASA (1976) *Landsat Data Users' Handbook*. GSFC Document 76SDS-4258, NASA Goddard Space Flight Center.
- Norwood, V.T. & Lansing, J.C. (1983) Electro-optical imaging sensors. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 335–367. American Society of Photogrammetry, Falls Church, Virginia.
- Slater, P.N., ed. (1983) Photographic systems for remote sensing. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 231–291. American Society of Photogrammetry, Falls Church, Virginia.
- Vane, G. & Goetz, A.F.H. (1988) Terrestrial imaging spectroscopy. *Remote Sensing of Environment* **24**, 1–29.
- Vane, G. & Goetz, A.F.H. (1993) Terrestrial imaging spectrometry: current status, future trends. *Remote Sensing of Environment* **44**, 117–126.
- Vane, G., Green, R.O., Chrien, T.G., Enmark, H.T., Hansen, E.G. & Porter, W.M. (1993) The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment* **44**, 127–143.

4 Photogeology

Photographic images are the cheapest and most readily available remote-sensing data for geological interpretation. They may be acquired directly by cameras mounted on aircraft or satellites, or produced from digital image data. Images in this form, either black and white or colour, are available for all parts of the EMR spectrum accessible to remote-sensing systems. For simplicity, however, this chapter deals only with those displaying information from the visible and near-infrared regions. Analysis of longer wavelength data and those in digital form is covered in Chapters 5, 6 and 7. The simple photogeological principles outlined in this chapter form the basis for interpretation of any images, whatever their origin.

The majority of images used here are vertical black and white photographs from aircraft, together with some single-band photographic hard copy from the Landsat MSS and RBV systems. For many users, colour photographs are a luxury, but they do have certain advantages stemming from the human eye's better chromatic discrimination. A few natural and false-colour photographs and some discussion of their information content are incorporated in the chapter. Colour imagery is discussed at greater length in Chapter 5, however, because the results of computer processing of digital data are nearly always displayed in colour. Single photographs can be interpreted in two dimensions, indeed for Landsat images this is the only option, except for small strips of overlap between adjacent paths. However, stereoptic viewing of overlapping photographs adds the extra dimension of relief, which usually is exaggerated. As well as showing topography, stereoptic viewing permits the measurement of differences in altitude and the calculation of slope angles. These factors are of great importance both in highlighting geological features and enabling field sites and traverses to be located accurately. Where possible, therefore, the images in this chapter are in the form of stereopairs, set up for viewing with a pocket lens stereoscope. Appendix A contains some discussion of stereometry.

Interpretation of photographs for any purpose relies on several basic characteristics of surface. These are tone, texture, pattern, shape, context and scale. They are all more or less qualitative attributes, and their use is very much a matter of experience and personal bias.

Photographic tone refers to the colour or relative brightness of parts of the surface making up a scene. In a black and white photograph it is expressed as different shades or levels of grey. Without tonal variations the shapes, patterns, textures and context of objects could not be seen. As it is related to the reflectance properties of surface materials, tone depends on the part of the spectrum covered by the imaging system and the composition of the surface itself. Tone also may be affected by processing and printing, and by the illumination conditions too. For example, the same surface will exhibit different tones on slopes facing the Sun, those facing away and those in shadow. As a result the absolute tone is of less use in interpretation than the relative tonal differences between different objects. For instance, a stand of deciduous trees in full leaf will have a lighter tone than conifers, irrespective of lighting conditions or the method of printing. Except under unusual circumstances in some arid or semiarid climates, rocks are generally masked by soil and vegetation. Even when cropping out at the surface they are frequently veneered with lichens, mosses or inorganic patinas. As a result rocks only very rarely show their true colour or tone. None the less tonal differences do play an important role in lithological discrimination on photographs (Section 4.2).

Texture is a combination of the magnitude and frequency of tonal change on an image. It is produced by the aggregate effect of all the many small features that make up a particular area of surface. The texture of a forest may be made up of the effect of branches, individual trees and stands of individual species. The scale and resolution of the image, however, will determine which of these dominates texture. On a Landsat MSS image with 80 m resolution only stands of trees will affect texture, but on a low-altitude aerial photograph all the attributes of the forest except the leaves will contribute. Likewise, the texture associated with a particular rock type will change with scale and resolution.

Patterns on an image result from the spatial arrangement of the different tones and textures that make it up. They may be arrangements of vegetation, topographic features, drainage channels or different types of rock, which can be traced discontinuously over a significant area. The patterns used most commonly are those relating to drainage networks (Section 4.1.1), which are often

Many of the figures in Chapter 4 are stereopairs of aerial photographs set up for viewing with a standard lens stereoscope. If you do not have stereoscope, or find difficulty in achieving stereoptic 'fusion', you will find all the stereopairs as anaglyphs in the *Image Interpretation in Geology* directory on the CD, within a separate folder. Use the anaglyph viewer packed with the CD to view them in any graphics package, such as Adobe Photoshop or MS Photo Editor.

related to underlying rock type or geological structure. Sometimes patterns on an image delineate specific shapes that can be related to familiar geological features, such as folds, igneous intrusions, faults, deltas and volcanic cones. Discerning and interpreting familiar shapes relies on the interpreter's powers of observation and experience.

Precisely how particular tones, textures, patterns and shapes are interpreted depends to a marked degree on their location relative to known attributes of the terrain under scrutiny. Their context is important. A U-shaped valley in the Sahara Desert is unlikely to have been formed by glacial action, and most probably results from water erosion during sudden floods. Similarly, a circular feature in an area of metamorphic rocks is unlikely to be a salt plug. By the same token, the scale of a feature may preclude some interpretations. A depression 40 km across bounded by arcuate faults is unlikely to be a karstic feature, but may well be a product of volcanic activity.

The criteria outlined in the following sections depend on various combinations of these basic characteristics of images. In photointerpretation they are usually combined together with topographic relief expressed in stereoptic viewing, where possible. Obviously, tone, texture, pattern and shape, because they depend partly on illumination conditions and sometimes on vegetation cover, may change depending on the time of day and year of the imagery. In many cases there is an optimum time of day or a best season when geologically related features show up most clearly. Very subtle slope changes related to geology in a flat terrain may be highlighted by low Sun angles or by remnants of snow or frost occupying depressions. The same viewing conditions in a rugged terrain may degrade the interpretability of a scene because of too much shadow or snow relics, so that high Sun angles in summer give the best results. As sedimentary rocks vary in their porosity, and because water content affects their reflective properties, it may be better to use images that were taken after rains rather than those from the dry season. However, in an intensely farmed area the post-wet season period may be characterized by the influence of dense crop canopies. These mask reflectance from rock and soil and disrupt natural patterns through the superimposition of field boundaries. In some parts of the world dry season images are best because the only thriving vegetation is natural and may be related to the underlying geology.

4.1 Destructional landforms

The rock-sculpting activities of the various agents of weathering and erosion in different climatic zones determine both the textures and patterns of the surface. It is from these features that interpretations of bedrock geology

are made. The transport and eventual deposition of erosional debris has a contrary effect. It masks features relating to the hard rocks buried by it. These two fundamental aspects of the Earth's surface processes produce two basic types of landform. Destructional landforms lay bare to some degree the structure and composition of consolidated and lithified rock, and indicate the dominant environmental influences over the development of landscape. Constructional landforms result from a superficial veneer of relatively recent rocks and sediments. Their surface attributes also give clues to environmental processes as well as to the nature of the veneers themselves.

In this section the emphasis is on destructional landforms, as they lead into the criteria for 'hard rock' interpretation covered in Sections 4.2–4.4. Constructional landforms and their associated superficial deposits are considered in Section 4.5.

Different rock types respond to weathering and erosion in different ways. This depends on many factors. The most important are the composition and degree of lithification of the rocks themselves, together with the type of weathering and erosion. These last two factors are mainly determined by climate, and sometimes by altitude. A region in the tropics may have a hot, humid climate near sea level, but mountainous areas experience conditions somewhat similar to those of temperate latitudes. Likewise, mountains with a moderate rainfall may flank an arid or semiarid desert.

An important criterion in geological interpretation of images is the degree to which different rock types resist erosion. A sandstone strongly cemented with an inert compound such as silica will be resistant under virtually any conditions. A rock containing reactive minerals such as feldspars or carbonates, however, will soon succumb to the agents of chemical weathering that predominate in hot humid climates. The same rock may respond in the opposite fashion in an arid or cold climate where chemical weathering is less significant. A rock's resistance to erosion is also relative. A basaltic dyke intruding soft mudstones may give rise to a prominent ridge, but where it cuts gneisses it may have a relatively low resistance and occupy a narrow trough.

There are four important types of erosion that can attack the products of weathering—those associated with flowing water, wind, wave action and moving ice. Marine and wind erosion can produce distinct destructional landforms, but they are of minor importance in image interpretation. Marine erosion is restricted, for obvious reasons, and most geological investigations are conducted well away from the sea. Aeolian erosion has a relatively minor effect at the scale normally used in remote sensing. The major sculpting agents are flowing water and ice. Water dominates, even in the most arid of modern environments, either because the main landforms were produced during more humid episodes or

water continues to flow at the surface after intermittent but spectacular rainstorms.

4.1.1 Landforms and drainage

Except in strongly glaciated areas, drainage patterns conform to some degree to the regional slope of the terrain and to its underlying geological structure. Because of this it is often possible, even where outcrop is poor, to recognize the gross structural elements. The first step in using drainage, usually the most obvious element of a landscape, is to seek genetic connections between stream patterns and the disposition of rocks. This involves assessment of the geomorphological history of the area and the stage of development of landscape dissection. It is not possible to catalogue all possible combinations of rock type, structure and stream development, or to outline all the many methods of describing drainage. This section uses a simple classification based on the direction of stream courses relative to geological controls.

A channel that has developed on some initial slope, and which continues to follow its original direction is said to be consequent. The original slope may have resulted from regional tilting of an earlier peneplaned surface, or may result directly from the direction of dip of a stratigraphic unit. The point to remember is that the term relates only to the topographic, and not any structural aspect of a consequent stream's origin. Streams that clearly flow down the dip of a particular stratum or lithological boundary are termed resequent streams. They have developed on a surface already stripped by earlier, consequent streams and are structurally controlled.

Streams propagate themselves and develop tributaries by headward erosion. They tend to follow the lines of least resistance. On a surface underlain by dipping strata that resist weathering and erosion differently, some will follow lines determined by the least resistant layers. These are subsequent streams. By the same token, subsequent streams can follow and etch out the lines of fractured rock developed along faults. Consequently the strike of strata and fault lines, so important in unraveling geological structure, are most likely to be reflected in subsequent stream courses.

Continued downward and headward erosion eventually begins to etch out the basic geological fabric of an area, in the form of *cuestas*, valleys and other topographic features. Slopes are produced on the valley sides of original consequent and subsequent streams in concert with the underlying geological diversity. They may be in several directions. Streams that follow them are obsequent. In areas of dipping strata, slopes develop in subsequent stream valleys, which are in the opposite direction to the local dip. Small obsequent streams frequently develop on such slopes. Others may flow down the sides of major consequent or resequent valleys, with

a less obvious relationship to structure. By no means all streams conform to this simple classification. Most bear no obvious relationship to geology or former land surfaces. They are referred to as insequent streams.

The evolution of a landscape can be complex, particularly during protracted or episodic uplift and downward erosion. A stream system may begin by developing on a geologically simple surface of uniformly dipping strata, producing consequent, subsequent and other streams. This simple sequence may be merely a veneer unconformably above more complex older rocks. The older rocks eventually may be crossed by drainage that bears little resemblance to their structure, but is inherited from an earlier, higher level system. This results in a superimposed drainage system. Relics of it may persist even after prolonged erosion establishes the control of the deep structure, leaving many insequent streams. Figure 4.1 shows the main classes of stream in a diagrammatic way. Figure 4.2 is a stereopair of aerial photographs illustrating real examples and their use in interpretation.

In areas where tectonic activity resumes after the establishment of a drainage system, the old streams may continue in their original courses across the axes of uplift or folding to produce antecedent drainage. This can happen if the uplift is quite slow or if the streams involved are sufficiently powerful to erode through the rising crust fast enough. The best known examples of antecedent drainage are those of the Colorado River cutting through the Kaibab Uplift to form the Grand Canyon (Plate 4.1), and the Indus and Brahmaputra which flow southwards through the Himalaya, despite the phenomenally rapid rates of uplift in this tectonically active area. Active tectonic uplift in many cases is too rapid for antecedent drainage to develop, and streams may be reversed or ponded to form lakes.

Any cycle of stream erosion passes through various stages, which relate to how close the drainage system has adjusted to its erosional base-level. The famous geomorphologist W.M. Davis compared these stages to those in a human life, and his general terminology is still in common use today. A cycle begins when the base-level falls, either through uplift, tilting, fall in sea level, down-cutting of a major river, increase in stream flow or various permutations. Such a change is referred to as rejuvenation. In the initial stages the previous drainage system remains or a new one develops in depressions in a newly emerged land surface. The youthful stage is characterized by active erosion and the formation of destructive landforms. Streams flow in V-shaped valleys, and they almost completely occupy the valley floors. Maturity begins as the stream system begins to develop wide valleys as a result of the destruction of ridges and the beginning of deposition. Mature streams have reached a lengthwise profile in equilibrium with the base-level. Old age is characterized by the dominance of

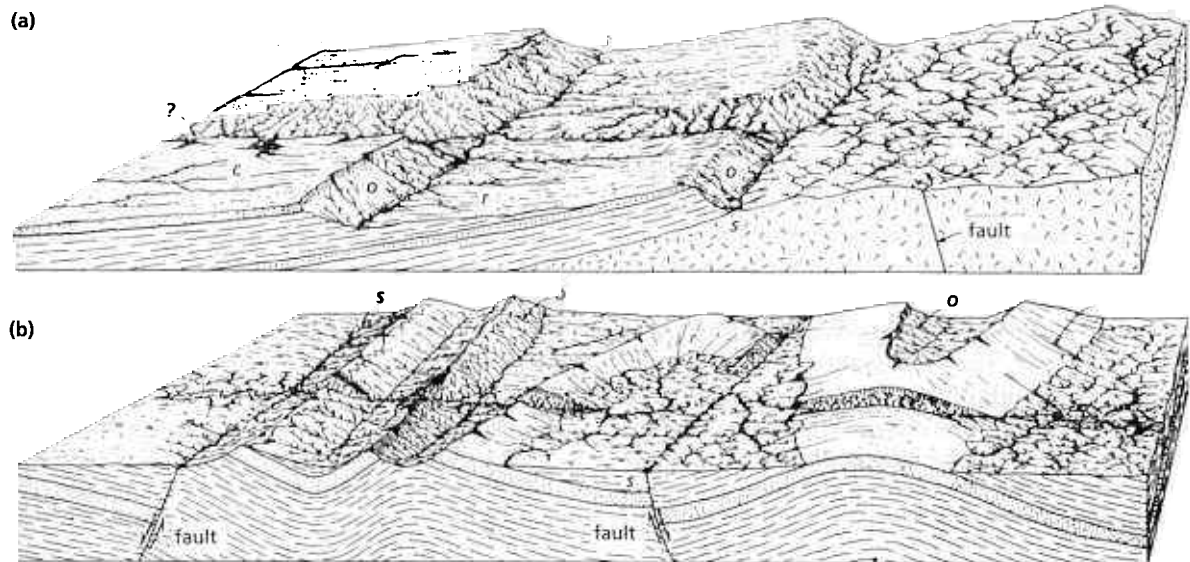


Fig. 4.1 These two diagrammatic sketches of how underlying geology may control the trajectories and patterns of surface drainage are from Miller's classic, but out of print, textbook *Photogeology* (1961). The letters c, s, r, o and i refer to consequent, subsequent, resequent, obsequent and insequent

streams, respectively. That marked '?' is possibly a consequent stream that developed upon a dipping planar surface, which has been incised into the underlying tilted strata during uplift. Reproduced with the permission of McGraw-Hill Books Inc.

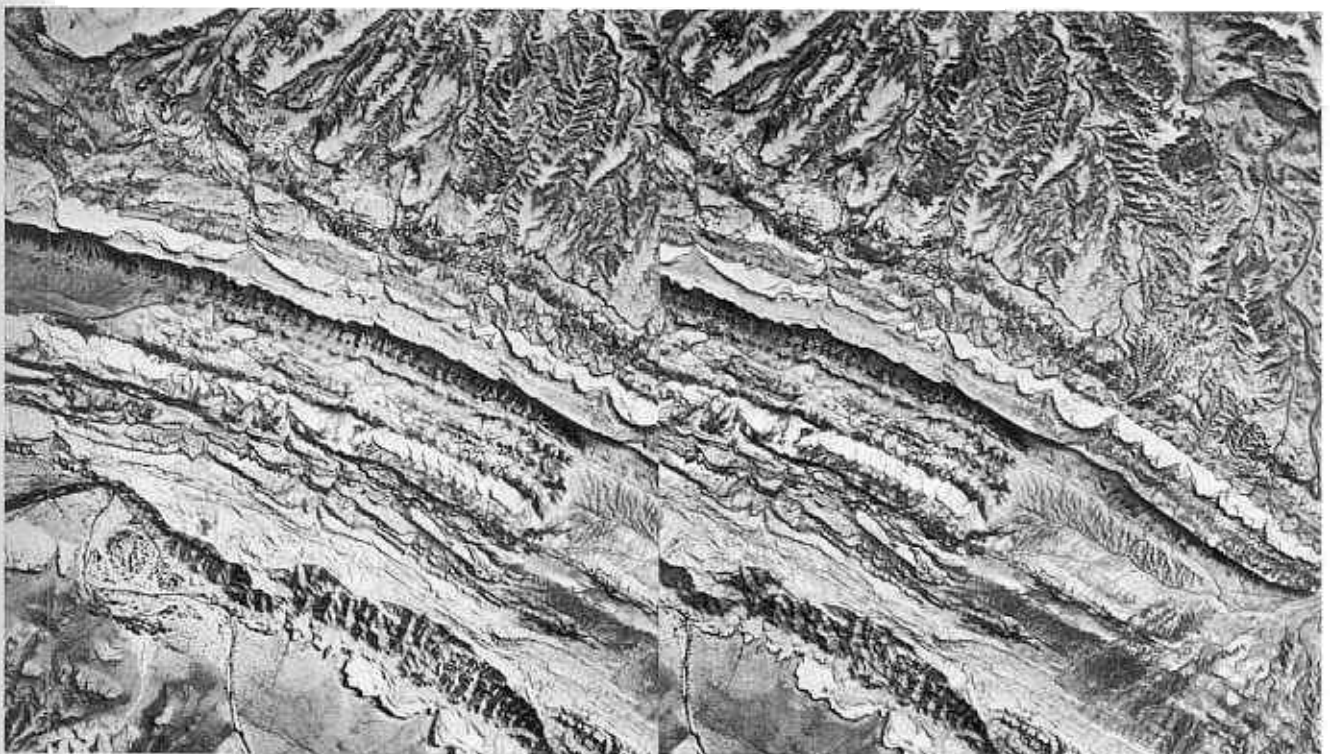


Fig. 4.2 Many of the dry stream channels in this semiarid part of Utah can be classified using the scheme in Fig. 4.1. A particularly important feature is the example of headward erosion in the major subsequent valley. Most of the gully profiles are sharp, which is characteristic of intermittent stream erosion. Scale 1 : 40 000. Source: ASCS UPRM-13 20 + 21. Note: Abbreviations in captions to figures using aerial photographs indicate their sources: ASCS—US Department

of Agriculture, Agricultural Stabilization and Conservation Service; CSL—Clyde Surveys Limited, UK; DNDE—Department of National Development and Energy, Australia; EMRC—Department of Energy, Mines and Resources, Canada; IGN—Institute Geographic National, Paris; NCIC—National Cartographic Information Centre, US Geological Survey; USGS—US Geological Survey. Where possible, photograph reference numbers are shown.

wide flat flood plains with meanders and redistribution of superficial sediments. The final result is a peneplain devoid of control by underlying geological structures. The mature and old-aged stages produce constructional landforms (Section 4.5).

How quickly an erosional cycle passes through these stages depends on a number of factors. The closer a stream is to the base-level, the faster it reaches equilibrium. More remote areas pass from the initial to youthful stages at rates dependant on headward erosion in the main streams. Easily eroded rock types will achieve maturity more quickly than resistant ones. In an area of varied lithologies youthful features may be mixed with those of maturity. The rate of erosion, determined by both volume and rate of flow, and by the erosive content of solids in streams, plays its part too. Fall in base-level is rarely instantaneous, and the rate of change also plays a role in the evolution of landscape. Episodic changes are common, and their interruption of the progress towards equilibrium can be marked by several levels in a drainage system, all characterized by youthful, mature and aged features. In such cases, the boundaries between levels can be marked by sudden changes in the longitudinal profiles of streams.

Adding to the complexity of landforms dominated by flowing water is the opposite of rejuvenation, when the base-level rises. This hastens the maturation of a drainage system and the onset of depositional features. Such aggradational episodes may indeed be interspersed with those of rejuvenation. Further consideration of these complex geomorphological processes is beyond the scope of this book, but are covered in many modern texts on physical geology.

The shapes of valleys in cross-section, which are often apparent in stereoptic views of photographs, can be important criteria in identifying underlying lithologies. Similarly, the drainage density can play an important discriminatory role (Section 4.2). However, both are strongly affected by climate. Surprisingly, drainage density in an arid area on any particular rock type is usually greater than one developed under humid conditions. Less obvious is the contrast in valley profiles between humid and arid terrains. Humid conditions encourage the rapid formation of thick soil profiles. Lubricated by water, these tend to creep slowly down-slope, thereby limiting slope angles and helping smooth the overall terrain. Under arid conditions neither soil nor lubrication is present to any marked degree. Consequently much steeper slopes to valley sides can be sustained, even in rocks that are not strongly resistant to erosion. Moreover, most streams in deserts are intermittent, and when they do flow it is often during storms. Under these conditions downward erosion is rapid but short-lived. It produces slot-like gorges along main channels, and, because torrential rainfall contributes

mainly to surface run-off, innumerable rills and gullies on the slopes. The soil and vegetation mantle in humid terrains slows surface run-off and encourages infiltration. Gullying is therefore more subdued, if present at all. Figures 4.2, 4.6, 4.7 and 4.8 show the contrasts between typical humid and arid terrains for these two important characteristics.

As implied earlier in this section, the patterns developed in a drainage system as it evolves contain important clues to geological structure. They can also make some contribution to distinguishing some rock types. These aspects are considered in Sections 4.2 and 4.4. Here the main pattern types are described and illustrated with some typical examples.

Drainage patterns of all types may be found on a variety of scales, from those defined by major river systems to those associated with rills and gullies. They help define the textural features of a scene, and depending on the spacing of the individual streams may be defined as coarse-, fine- or intermediate-textured patterns. The streams that define the patterns can be of any genetic type: consequent, insequent, subsequent and so forth.

Areas drained primarily by insequent streams show a lack of any preferential channel direction. The streams flow in irregular, branching courses with many random bends. Tributaries enter larger streams usually at an acute angle to the general direction of regional flow, itself determined by regional slope. The pattern produced is like the complex branching of a tree, and because of this it is termed dendritic (Fig. 4.3a). Dendritic drainage patterns are characteristic of terrains showing lithological, structural and topographic homogeneity.

In contrast, drainage may be dominated by a number of parallel or subparallel streams, indicating a unidirectional flow. For obvious reasons this is termed a parallel drainage pattern (Fig. 4.3b). Regular deviations of courses from this result in fan patterns, which reach their ultimate in a completely radial pattern, where streams all flow away from a single point (Fig. 4.3c). The opposite case, where streams flow radially towards a common point (Fig. 4.3d) is called centripetal drainage. Such simple patterns all derive from topographic features and cannot be assumed to reflect any particular structure or lithology, although this is often revealed by closer inspection.

Other categories of drainage pattern can be regarded as embellishments upon the last four. They include subsequent streams that have been controlled by lithological and structural weaknesses, and so give more geological information. Patterns dominated by many straight streams with a variety of trajectories may define rectangular or angular patterns, depending on their angular relationships (Fig. 4.3e). A particular variant of these two types, where short tributaries connect to long straight major streams produces a pattern reminiscent of trained vines

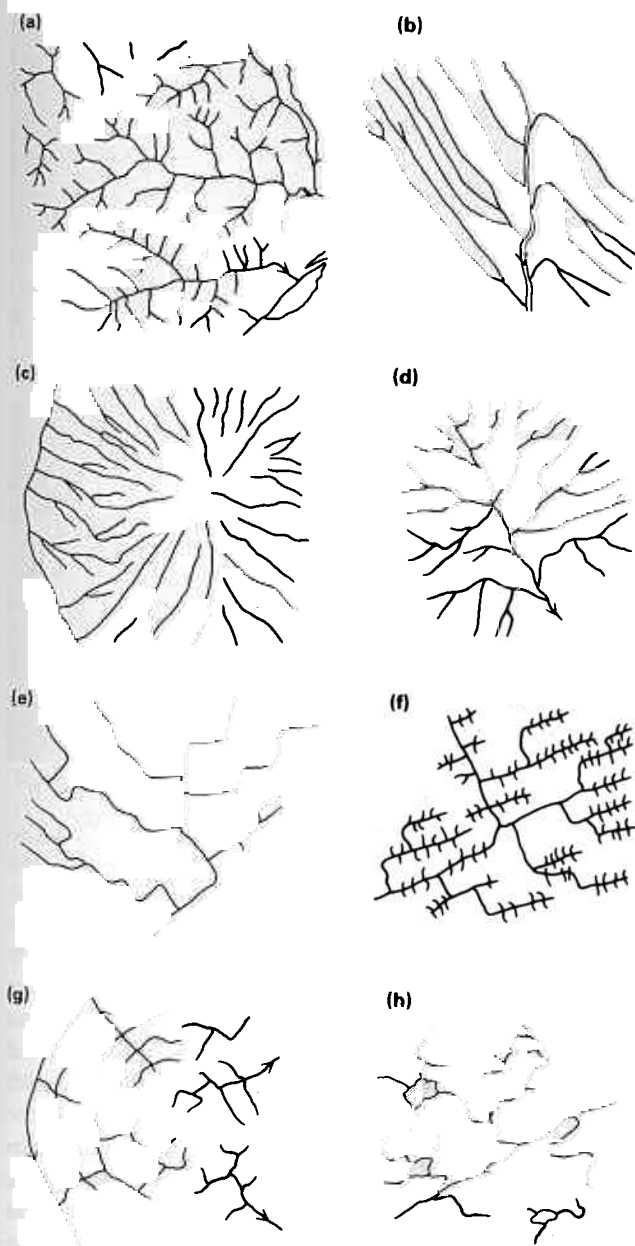


Fig. 4.3 Patterns formed by drainage networks can take on many distinct forms, often related to aspects of the underlying geology. A dendritic form is shown in (a), parallel in (b), radial in (c) and centripetal in (d). Minor variants on these four basic patterns include rectangular (e), trellis (f) and annular (g) systems. Areas with low, hummocky relief and few streams may develop deranged patterns (h). Many other subforms can be found but nearly all conform to this basic subdivision.

(Fig. 4.3f) referred to as a trellis pattern. Subsequent streams developed on a radial pattern result in a concentric series of valleys defining an annular pattern (Fig. 4.3g).

In reality, the several basic pattern types are rarely developed to perfection. Several generally combine together to give topologically complex systems, in which

specific elements may be picked out by the interpreter. Typically, an overall pattern or series of patterns may contain streams that do not conform to the overall picture. They are anomalous, and attract attention simply because they are different. Such drainage anomalies in an area of clearly defined geology may highlight geological subtleties, which otherwise would go unnoticed in a purely field study. On the other hand, in areas where geology is obscured because dissection is not well advanced or a thick blanket of superficial sediment masks the underlying structure, anomalies can give clues leading to geological revelations.

Whatever patterns are displayed on a photograph, it is always important to bear in mind that they reflect the full range of factors behind the development of landscape. Climatic, topographic, lithological, and structural controls all combine together with the stage reached in the evolution of the drainage system. In poorly exposed areas a crude idea of the structure can be appraised best simply by tracing the drainage patterns revealed by a photograph on to a transparent overlay. Sooner or later, however, interpretations have to be confirmed and extended by field studies.

4.1.2 Glacially eroded landforms

Moving masses of ice are the most powerful landscape-sculpting agents imaginable. Although velocities of ice flow are very slow, even in the most active valley glaciers, the enormous masses involved ensure that their energy is enormous. During the Pleistocene glaciations it is estimated that some of the continental ice sheets exceeded 3 km in thickness. The process of glacial erosion begins with the plucking of debris from the substrate, often weakened by freezing and thawing in advance of the ice front. The incorporation of this debris in ice provides the ice base with a renewable supply of abrasive material, which acts in the manner of a giant rasp. No rock type can resist for long the protracted passage of ice above it. The destructive landforms that result depend for their shapes and patterns on the type of glaciation involved. There are two main types—valley and continental glaciation.

Valley glaciation begins with the accumulation of permanent snow masses in depressions on mountain slopes. There it undergoes a transformation to ice through a series of stages. Once a certain mass of ice has accumulated it is able to flow downslope under its own gravitational potential energy. The climatic conditions suitable for ice accumulation ensure that exposed rock on peaks undergoes powerful physical weathering, dislodging angular fragments and thereby 'sharpening' the peaks. These fragments falling on to the ice, together with those plucked from the base of the ice, contribute to the abrasive capacity of the glacier. At its source the glacier

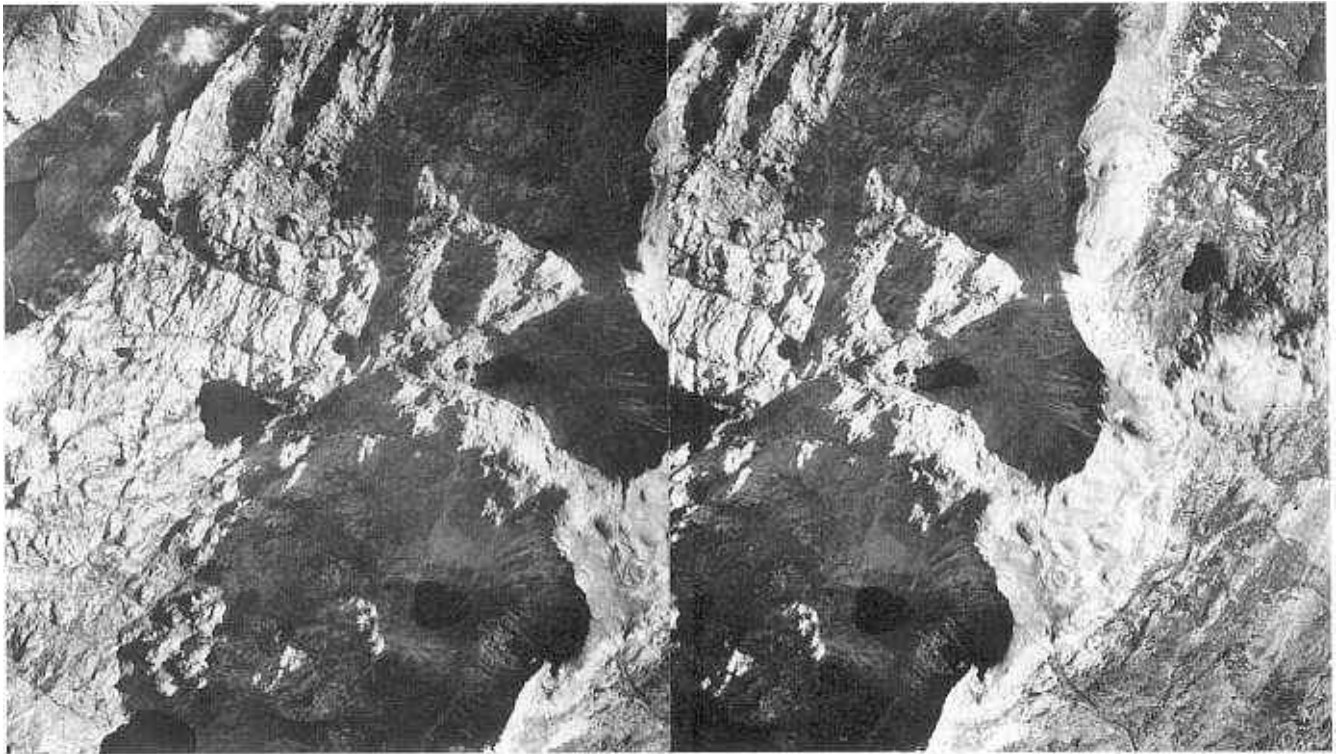


Fig. 4.4 The Highlands of Scotland contain many examples of the products of upland glacial erosion. This stereopair shows typical cirques, arêtes and hanging valleys, as well as the rugged bare rock surfaces resulting from subglacial plucking. Scale 1 : 25 000. Source: author.

carves a bowl-like depression or cirque by the local movement of ice through a vertical arc (Fig. 4.4). Once it has spilled over the lip of this bowl, the ice moves downslope, usually along a pre-existing valley.

Valleys are eventually remoulded to a U-shaped cross-section and straightened by the removal of interlocking spurs. Apart from their typical cross-section, glaciated valleys can be recognized easily by the abundance of truncated spurs. The other effect of glaciation is deepening of the earlier valleys, the amount depending on the energy of the glacier responsible. The pre-existing valley patterns determine to a marked degree the course of later glacial flow paths, which comprise major valleys with tributaries. The energy differences involved in the glacier system result in the courses of less powerful glaciers being less deeply eroded than those of major ice streams. This produces hanging valleys at the confluences (Fig. 4.4). Because valley glaciation is generally superimposed upon pre-existing stream and river courses, except in the Arctic and Antarctica, the remaining patterns conform to some extent to those of their precursors. The difference is that the minor valleys are wiped out and those remaining are grossly straightened. The net result is to preserve some of the indicators of gross geological structure, but to erase those indicating the details.

Examination of large-scale photographs of glacially eroded uplands enables some of the smaller features typical of ice erosion to be recognized. Among these are roches moutonnées and major ice striations (Figs 4.4 and 4.5). The typical shape of roches moutonnées, with smooth, rounded surfaces facing towards the former ice flow and steep, glacially plucked faces pointing down flow, enables estimates of ice flow directions to be made. Striations give similar information, but are ambiguous at these scales. In upland areas, ice-flow directions are generally obvious from the overall topography and the relative orientation of glaciated valleys, and these minor features are of more importance in glacial analysis of flat areas of continental glacial erosion.

During the Pleistocene huge ice sheets, originating from snow accumulation at high latitudes and to a lesser extent from ice-capped mountain ranges, spread southwards over large tracts of northern Eurasia and North America. They produced two main types of terrain, those dominated by glacial erosion and those where the products of erosion were deposited. The former are now the enormous rocky wastes of the Canadian, Baltic and Siberian Shields, from which most of the previous cover of Phanerozoic sediments was stripped. This revealed the underlying Precambrian basement. The areas dominated by glacial deposition occur further to the south where

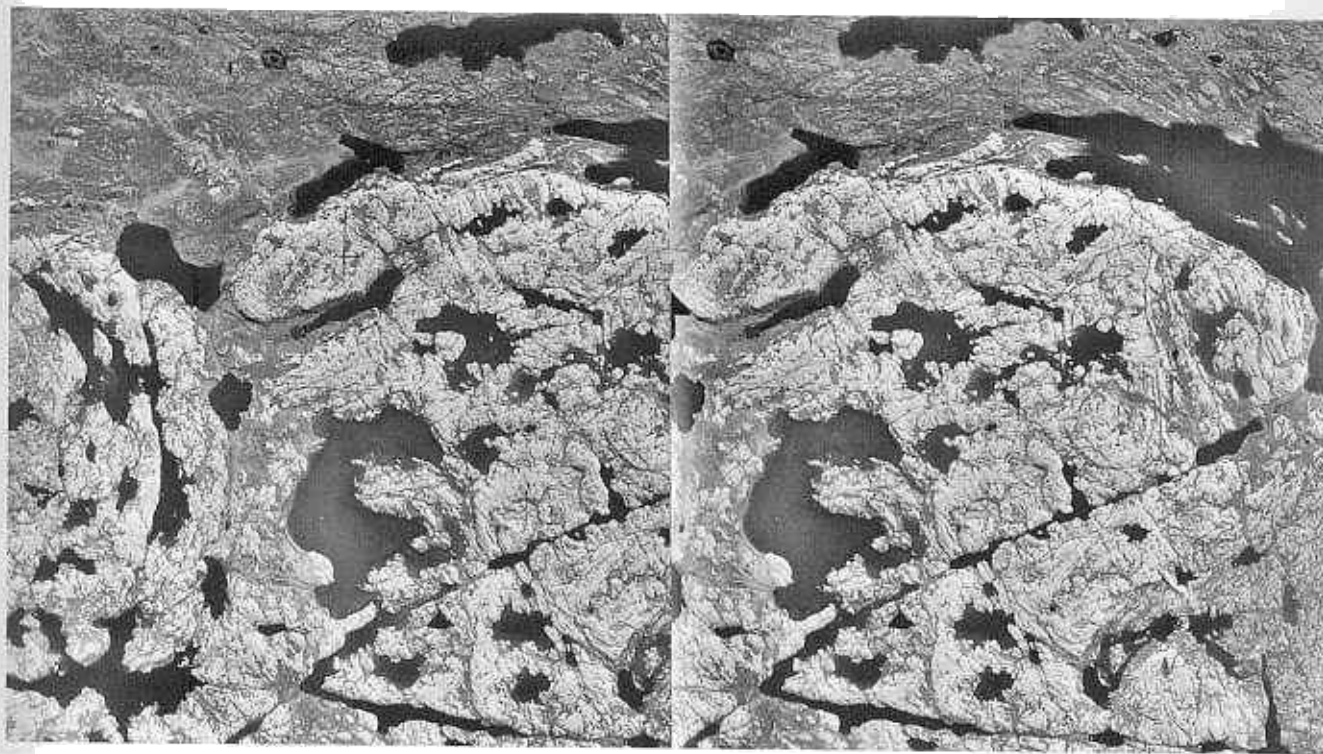


Fig. 4.5 Despite the hardness of the metamorphosed Precambrian rocks of the Canadian Shield they were unable to resist the erosive action of the vast ice sheets that drove over them during the Pleistocene. This stereopair shows the typical low relief of such areas of continental glaciation. At top right can be seen grooves and ridges reflecting the ice flow direction. There are two main rock types exposed, pale granite and

darker metavolcanic rocks, with the granite having resisted glaciation better. The intricate jointing in the granite, its intrusive contact and basic dykes cutting it have been selectively exploited by glacial action resulting in linear, vegetation-filled depressions and narrow lakes. Scale 1 : 50 000. Source: EMRC A13781 52 + 53. Copyright HM the Queen in Right of Canada (1953).

the impetus of the ice sheets waned. The forward movement of ice was balanced by the increased rate of melting at lower latitudes. As a result the ice sheets were relieved of their loads of debris. The typical features of these depositional areas are considered in Section 4.5. Here a few comments about the erosional features of large, uniform ice sheets are required.

Unlike upland glaciers, for which erosional features are restricted to more or less narrow valleys, continental ice sheets have a uniform and dramatic effect on the areas that they have stripped bare. Although in a general sense they brook no obstacles and literally raze them to the ground, detailed examination reveals effects of great significance to the geologist. The all-pervading erosion by the debris-loaded ice base emphasizes every conceivable heterogeneity in the rocks over which it passes. The softer rocks and zones of structural weakness are etched out, sometimes completely, leaving the more resistant rocks as positive features. Pre-existing drainage patterns are obliterated, so that when the ice retreats new drainage is controlled by the intricacies of the newly scoured surface (Fig. 4.5). These patterns can be long-lived, simply because the grinding flat of the land surface reduces stream energy to a minimum. However, the

transient but massive load of the former ice sheets was sufficient to depress the crust hundreds of metres below its original level. Slow post-glacial isostatic rebound, at different rates and to different extents depending on the amount of glacial depression, has imparted regional slope to the shields. These ensure that the drainage does have an overall orientation, albeit tortuous.

4.1.3 Other destructional landforms

Marine erosion is capable of producing landforms that to some degree are controlled by the underlying geology. They are restricted, however, to a narrow coastal swath, and in general are obvious only on large-scale photographs. In some cases, a rise in sea level and flooding of a landscape formed by other processes can reveal gross geological elements from the patterns of the islands left above sea level. Where geological interpretation of coastal areas is demanded, the reader is referred to existing texts on coastal geomorphology.

Erosion by wind in deserts is indeed an important process, but its products are generally visible only in the field. The reason for its relatively minor importance to the image interpreter is that wind energies are low compared

with those of flowing water. Because of this only a small load of abrasive particles can be carried. Wind erosion is rarely powerful enough nor at a sufficient rate to overwhelm the dominant effects of stream erosion in deserts, even though streams flow only intermittently.

4.2 The recognition of rock types

In order to recognize or distinguish between different lithologies on a photograph the geologist needs two prerequisites—a sufficient correlation between geology and landform and a combination of experience, patience, perception and ingenuity. The attributes of photographic tone, texture, patterns, context and scale need to be welded together with a knowledge of how different rock types respond in different climates to weathering and erosion.

In this Section the basic division is into sedimentary rocks, extrusive and intrusive igneous rocks and metamorphic rocks. To some extent the treatment pre-empt later Sections by introducing some of the stratigraphic and structural criteria required in rock discrimination. Most of the distinguishing criteria, particularly those relating to textures and patterns, are equally useful in the interpretation of images of other parts of the EMR spectrum covered in later chapters. As many rock types have different appearances under different climatic conditions, wherever possible they are illustrated with several images. It would be useful to give examples of the kinds of appearance most likely to be encountered by the interpreter—unspectacular and poorly expressed on the image. However, the world is a diverse place, this is an introduction and space is limited. The illustrations here are the best that could be found.

There is no rigid routine for the recognition and discrimination of lithologies in photogeological interpretation. However, a systematic approach of one kind or another is essential. It must be backed up by note taking and annotation of the images, so that a permanent record is kept and experience is built up in a way that can be referred to at any future time. As a guide to this approach the following general scheme is recommended.

- 1 The climatic environment has to be assessed—humid or arid, boreal, temperate or tropical.
- 2 Is the vegetation, if any, natural or agricultural?
- 3 The erosional environment must be assessed in terms of its energy, its stage of development and the relative contributions of stream, glacier and wind erosion estimated.
- 4 The area should be divided into parts that are dominated by superficial deposits and those probably having bedrock at or close to the surface.
- 5 Areas of outcrop or near-surface bedrock should be examined for signs of lithological banding or its apparent absence, and distinguished accordingly.

The next set of steps depends on the systematic description of the probable areas of bedrock. If possible, the terrain should be examined closely for any obvious boundaries between different types of surface. After a preliminary division of contrasted surface types, each can be described using the following key headings.

- 1 The topography should be described, in particular noting the relative resistance to erosion of the unit and any anomalous features.
- 2 What is the drainage pattern and what is the spacing between individual channels?
- 3 If possible the cross-sectional shape of the smallest drainage features—gullies—should be described.
- 4 The vegetation cover and land use directly related to the unit must be noted.
- 5 Finally, if bare soil or rock outcrop is visible its photographic tone or colour compared with other units, and any textural features should be described.

4.2.1 Sedimentary rocks

Sedimentary rocks are usually stratified owing to variations in the conditions of their deposition. As a result they typically display a banded appearance on photographs, although the width of bands depends on their attitude as well as on the thickness of individual strata. As volcanic lava flows and pyroclastic rocks, some intrusive igneous rocks and most metamorphic rocks also show layering, there is considerable scope for confusion. Avoiding this relies on the geological context of the layered rocks being examined. In most cases this is obvious from previous knowledge of the study area. In relatively unknown areas the context must be assessed first by looking for signs of volcanoes, cross-cutting contacts and highly complex structures. In their absence, layered rocks can be assumed to be sedimentary until field study proves the contrary.

The topographic expression of sediments depends on a variety of factors. Principally it is determined by the attitude of bedding, ranging from horizontal to vertical. This can result in a series of landforms from which estimates of dip and strike can be made, but this is considered in detail in Section 4.3. Topography is also governed by resistance to erosion, which is owed to a variety of factors in its own right.

Sediments can exhibit various stages of compaction and cementation. For example, a compact but uncemented sandstone is less resistant to some types of erosion than one that is completely cemented. The porosity and permeability play a role too. A permeable rock will allow water to seep into it thereby lessening the effect of erosion by surface run-off compared with that on an impermeable rock. The main weak zones in all sediments are bedding surfaces. Their spacing makes an important contribution to relative resistance. So a thinly bedded

rock is less resistant than one of the same composition with thick beds. For a sequence of equally well-cemented sediments, those comprised of small particles will erode more easily than the coarser ones. This is because the energy to dislodge small grains is less than that for large ones. Finally, rock composition is a contributory factor. As well as the hydraulic action of moving fluids, their content of particulate and dissolved matter helps abrade and dissolve rocks. Rocks such as sandstone and siltstone composed dominantly of quartz, which is both hard and chemically inert, resist both mechanical and chemical erosion. Mudstones are composed mainly of soft clay minerals, which although inert are easily abraded. Rocks containing a high percentage of soluble or reactive compounds, such as carbonates or various salts, respond easily to the effects of mildly acid rainwater containing dissolved carbon dioxide or organic acids.

The other morphological features of sediments, their associated drainage patterns, textures and gully shapes, depend on exactly the same fundamental factors as resistance. Vegetation cover and outcrop tone or colour are controlled by more subtle factors, and are described for each main sediment type. Only three main groups of sediment need be considered in photogeology: coarse clastic rocks such as conglomerates and sandstones, fine clastic rocks such as siltstones and mudstones, and biochemical and chemical sediments including limestones and evaporites.

Sedimentary sequences are nearly always composed of a variety of interbedded lithologies. The differences in their resistance to erosion impart a distinctive grain of parallel positive and negative features to the terrain. The appearance or bedding as patterns or textures depends on the attitude of the bedding.

Horizontal strata containing resistant units are generally expressed by distinctly stepped topography, especially when the resistors are more than 5 m thick. Each step is controlled by a resistant unit and separated by less steep slopes underlain by soft rocks. In arid climates, resistors frequently form plateaux incised by major streams flowing in steep-sided canyons with stepped walls. Headward erosion of tributaries can isolate buttes and mesas from the main plateau (Fig. 4.6a). Soil formation and its movement downslope under humid conditions mutes such topographic expression (Fig. 4.6b). The lack of original slopes in horizontally bedded terrain tends to encourage the development of dendritic drainage patterns. This, together with the tendency for major units to crop out parallel to contours results in complex patterns and textures (Fig. 4.6). Such complexities are often paralleled by vegetation communities, which may differ from one lithology to another.

Tilted sediments that have been incised by stream erosion provide the conditions whereby consequent, sub-sequent and other stream classes can develop. The full

exploitation of differences in resistance can then leave resistors as ridges of various kinds separated by parallel subsequent valleys in the less resistant units. The net result is a gross texture of more or less parallel elements of the landscape (Fig. 4.2), characterized by some form of trellis drainage pattern. The degree of parallelism becomes more nearly perfect the steeper the dip of the beds. In many humid terrains the ridges are wooded, whereas valleys are farmed.

Coarse Clastic Sediments

Sandstones and conglomerates are composed mainly of quartz grains, with varying proportions of other resistant minerals such as feldspars and rock fragments. Their degree and nature of cementation varies considerably, and so does their resistance to erosion. When interbedded with equally cemented finer clastic rocks, however, they generally form positive features, often forming cliffs and steep escarpments. Cements based on iron compounds or silica are particularly strong, but carbonate cements are prone to attack by acids in rainwater.

In arid climates, quartz grains dislodged by weathering are easily removed by wind. Sandstones and conglomerates therefore tend to display bare surfaces under these conditions (Fig. 4.7). Except in tropical rain forest, where laterites may develop, coarse clastic rocks in humid climates rarely develop deep, fertile soils. Their permeability makes them too well-drained to support extensive grass cover. As a result vegetation is relatively sparse and often dominated by trees with deep root systems (Fig. 4.7).

As many sandstones are very permeable, rainfall infiltrates them, thereby reducing surface run-off. This helps ensure that they have relatively wide-spaced drainage textures (Fig. 4.7b). Gullies are relatively rare and small streams tend to develop V-shaped cross-sections owing to the well-supported nature of sandstone's granular structure. Being unable to deform except by brittle dislocation, one of the most distinctive features of coarse clastic sediments is their abundance of joints, often in several tectonically related directions. The zones of weakness associated with jointing exert a marked control on the finer structure of drainage on sandstones, which may take on a prominent rectangular or angular pattern (Fig. 4.7a). Cliffs and cuestas in sandstone are frequently joint-controlled.

The bareness of sandstone outcrops in arid areas, combined with their well-drained nature, usually permits them to show their true photographic tone or colour, unaffected by near-surface pore water. Silica- and carbonate-cemented sandstones are usually pale, but those containing iron compounds can be any shade of yellow, orange, brown or red. In panchromatic black and white photographs red sandstones appear dark. The frequently

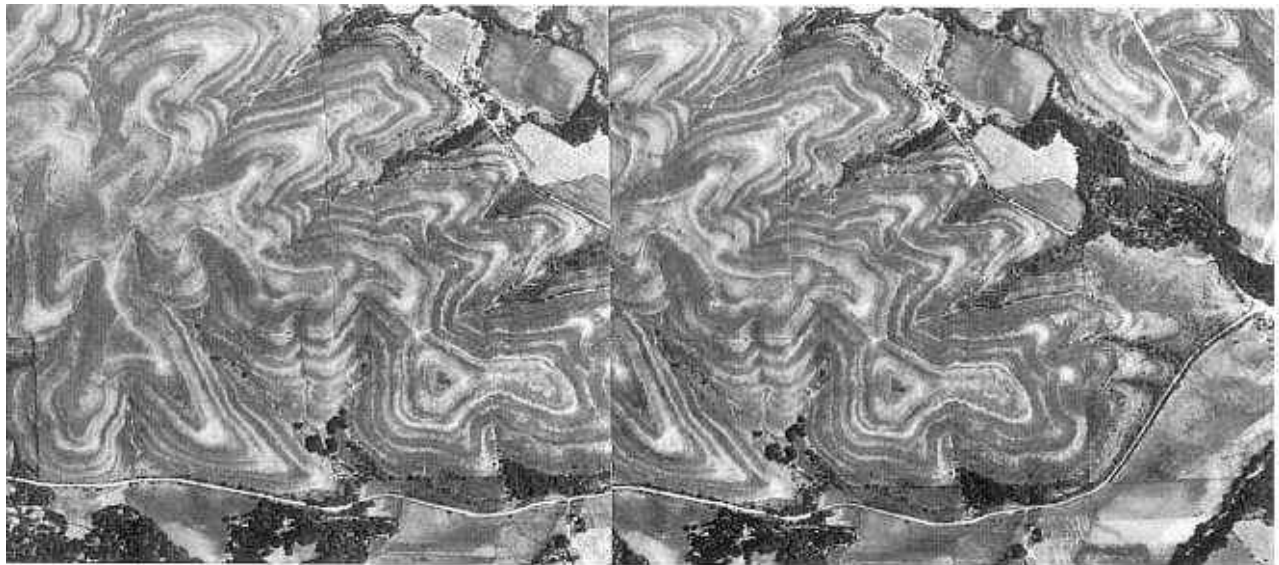
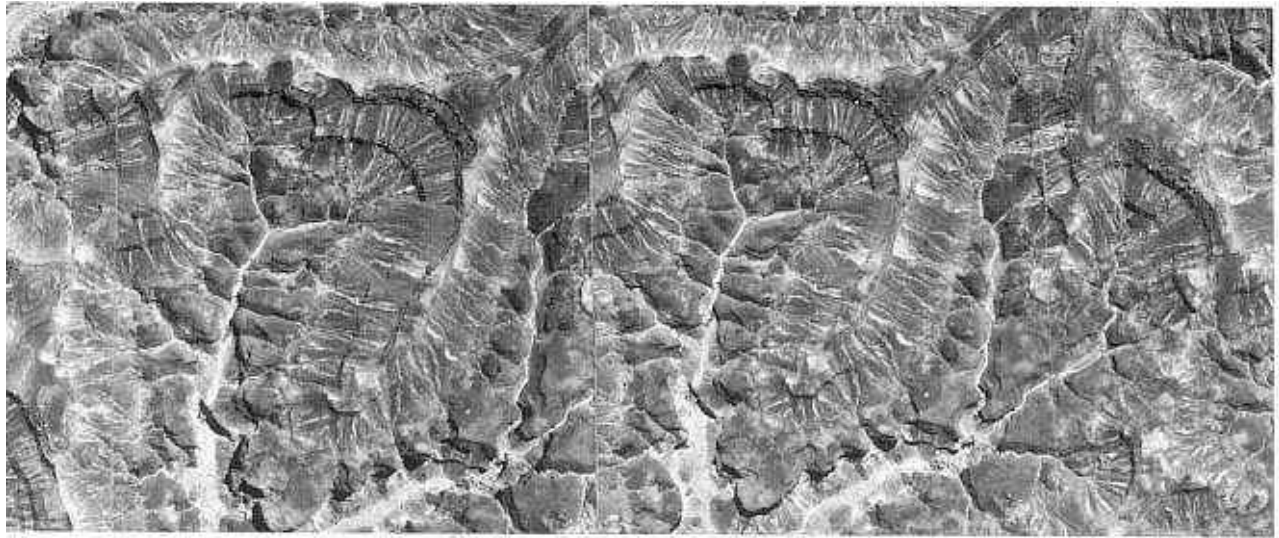


Fig. 4.6 Both stereopairs show the most typical feature topography developed upon horizontally layered outcrops that closely parallel topographic contours. In arid terrains (a) the intermittent violent erosion develops one-sided gullies and valleys. Variable resistance to erosion, as case by sandstones and siltstones, leads to sudden changes in slope angle and step-like features on the valley floor. One noteworthy feature is the restriction of bushes

to one level in the succession. This probably is a boundary between water-carrying sandstone and underlying permeable siltstone. In a more humid climate such as in eastern Kansas (b) the topography is more muted. The spectacular 'candy-stripe' appearance results from bedded white limestones and grey shales of Permian age. Scales: (a) 1 : 20 000 and (b) 1 : 20 000. Sources: (a) CSL Series B 3 + 4 (b) NCIC Kans.2 A + B.

The forest cover of sandstones in humid climates also gives them a dark colour on panchromatic photographs.

Clastic Sediments

Sandstones and siltstones contain a high proportion of clay minerals together with varying amounts of quartz. They represent depositional environments with low energies. In many cases they contain fine laminations between 1 mm and a few centimetres thick. Because of this such rocks

usually show a strong fissility parallel to bedding, which is accentuated by the reorientation of flaky clay minerals during lithification. It is this fissile nature which characterizes shale, a term that is commonly applied to fissile mudstones and siltstones. Fine-grained clastic sediments may comprise a variety of grain sizes, some in graded units, and show bedding on a scale from millimetres to 1 m. However, except in detailed stratigraphic work they are usually lumped together as homogeneous and may be up to hundreds of metres in thickness.

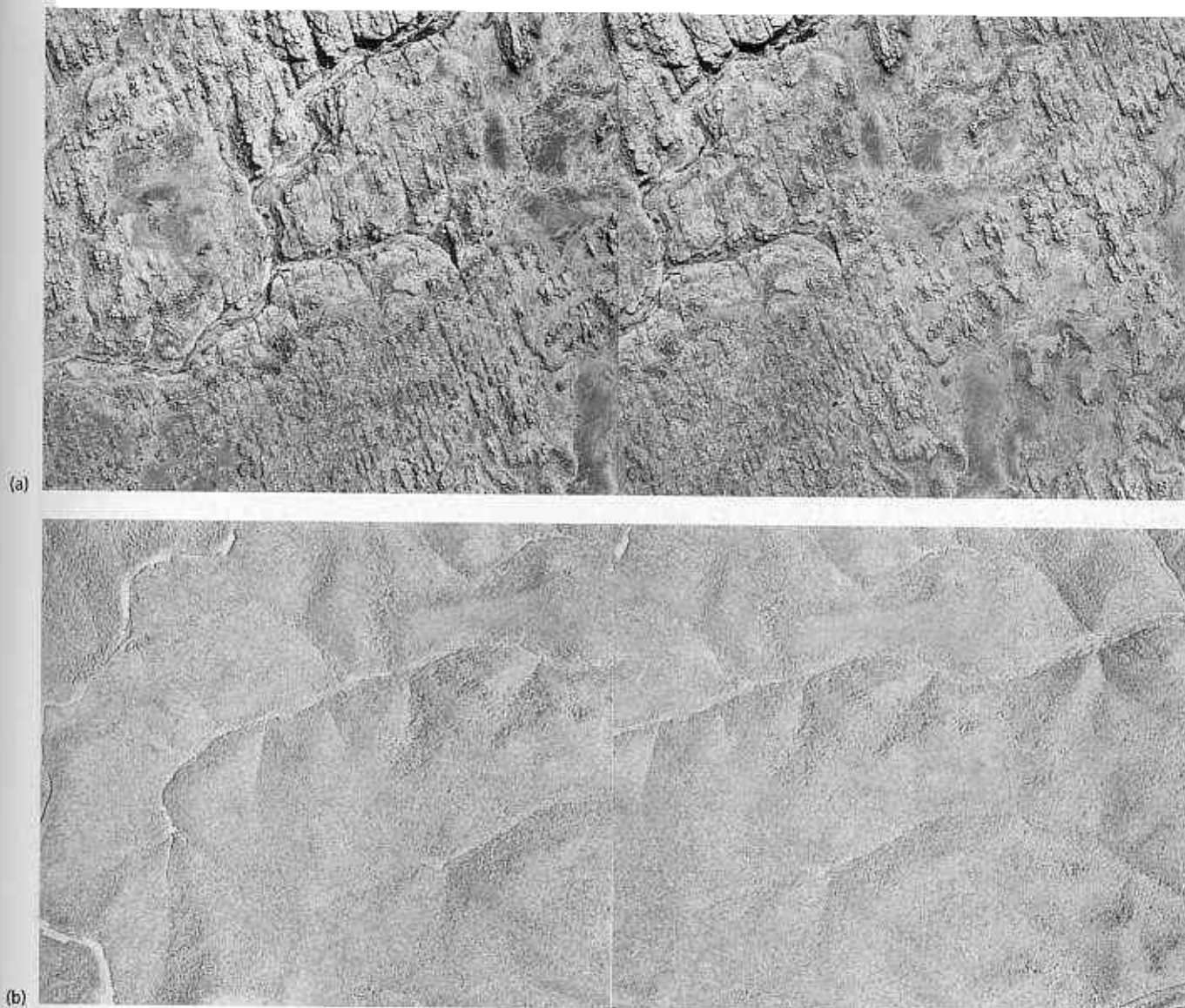


Fig. 4.7 (a) Sandstones in an arid climate, in this case in Utah, are typically bare and display their internal regular jointing very well. Their excellent internal drainage reduces the number of streams at the surface. High resistance to erosion results in a rugged surface. (b) Sandstone in humid climates develops very poor soil on which the dominant natural vegetation is woodland as in this part of Pennsylvania. The

main clue to the presence of sandstone is the wide spacing of drainage and the roundness of topography resulting from its often weak cement and susceptibility to chemical weathering. Compared with fine clastic sediments it is resistant to physical weathering, and is not prone to solution like carbonates.

Scales: (a) 1 : 20 000 (b) 1 : 33 000. Sources: (a) USGS GS-WI-12-146 + 147 (b) ASCS APL-4 V 98 + 99.

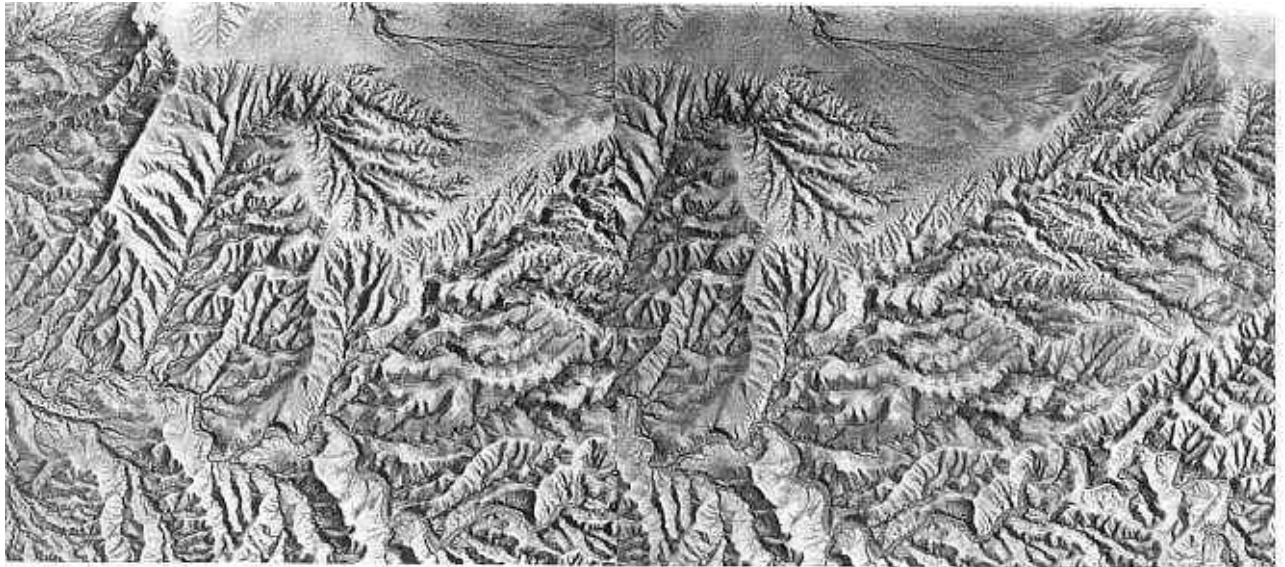
Cements can be of various compositions, but even when totally uncemented, rocks with a high clay content have good cohesion owing to the electrostatic charges on the clay mineral particles. The poor permeability of fine clastic rocks, however, means that in an unconsolidated sequence of fine and coarse sediments rainfall infiltration is at a minimum on the fine rocks and surface run-off has more chance of eroding them than the coarser materials. Irrespective of degree of cementation and climate, therefore, fine clastic rocks are more prone to

erosion than coarse clastic rocks. They tend to be poor resistors and form negative topographic features.

As well as promoting relatively rapid erosion, high run-off results in the development of closely spaced drainage textures. Clay-rich rocks are more able to deform plastically than sandstones and as a result tend not to show prominent joint systems, although they may have finely spaced joints as a result of shrinkage. This encourages the development of randomly orientated streams in horizontal strata and an overall dendritic drainage pattern.

In dipping sequences, fine clastic units often display a dendritic element within an overall trellis pattern of consequent and subsequent streams. The low resistance to erosion, particularly when cement has been dissolved or weakened, together with the continual exposure to surface runoff means that gullies in fine clastics soon develop. In an arid climate gullies form during short-lived heavy rainfall and aggregate to form a minutely dissected terrain, gully walls retaining steep angles. The continuous

runoff on fine clastic rocks in humid climates encourages the degradation of gully sides, the formation of less closely spaced drainage and a rounded topography. Any rock with a high clay content is likely to contain a higher complement of plant nutrients than those composed mainly of silica or carbonates. These are released for use by plants during weathering and soil formation. These soils are not only nutrient rich, but are able to retain nutrients and high moisture contents because of the



1.8 Gently dipping mudstones and siltstones typically develop dendritic and closely spaced drainage patterns because of their poor internal drainage. In arid areas, such as (a), individual gullies have a sharp V-shaped cross-section despite the low resistance to erosion of fine clastic sediments. Paradoxically, drainage on shales in humid climates (b) is often more widely spaced than under dry conditions. As shown clearly by the stereopair of part of

Arizona, the smallest channels have vanished as a result of the absence of lateral erosion during continuous stream flow. Gully profiles are smooth and shallow. The high content of nutrients in shales together with their poor internal drainage helps ensure that they have a luxuriant natural vegetation cover. Scales (a) 1 : 27 000 and (b) 1 : 43 000. Sources: USGS GS-WI-36 68 + 69 and (b) USGS GS-AZ-2119 + 120.

structure of clay minerals. In humid climates they therefore support dense vegetation and are often intensely farmed. In arid climates the converse applies. Unlike coarse clastic rocks they do not receive much infiltrated rainfall and are consequently more barren.

Fine clastic rocks can form under a wide range of conditions and may have very diverse chemistries, varying on a scale of metres. Iron, the main colouring agent in rocks, can be chemically bound in the clay minerals or exist in the cement. The low depositional energies of many mudstones encourages reducing conditions during deposition. This often results in a high proportion of iron sulphides in the cement. Weathering of sulphides produces a complex of iron oxides and hydroxides, collectively known as limonite, which can impart strong yellow, orange, brown and red coloration to outcrops. Unless they are composed of pure clay minerals such as kaolinite, fine clastic rocks generally show up as relatively dark surfaces on photographs. Provided that resolution is adequate, their heterogeneity may be revealed as fine colour or grey tone banding. Because of dense vegetation cover and ease of erosion, fine clastic rocks rarely crop out over large areas in humid climates and must be distinguished on grounds other than photographic tone.

Figure 4.8 shows the typical appearance of horizontally bedded fine clastic sediments in humid and arid climates.

Chemically Precipitated Sediments

There are two main categories of chemical sediments, those precipitated by evaporation of restricted bodies of water—highly soluble halite (NaCl) and other salts—and those precipitated mainly as the hard parts of organisms—calcite and dolomite, the carbonates of calcium and calcium with magnesium—which form various limestones.

Being highly soluble in water, evaporites rarely crop out in humid terrains. However, their presence in a sequence may be detected by the haphazard presence of more or less circular depressions where collapse into solution cavities has occurred. Where the evaporites are in the form of salt diapirs, such collapse structures may be arranged in distinctive patterns related to the shape of these gravity-driven structures. Near-surface salt may accumulate in soils as a result of capillary rise of saline solutions. As few plants are salt-tolerant this may give rise to barren patches in otherwise fertile ground.

Evaporites are crystalline solids, and being devoid of pores do not allow infiltration of rainwater. In arid climates, where they are attacked by only infrequent rain storms, salt bodies may crop out. They generally have distinctive textures composed of finely spaced gullies with sharp profiles (Fig. 4.9). Patches of residual

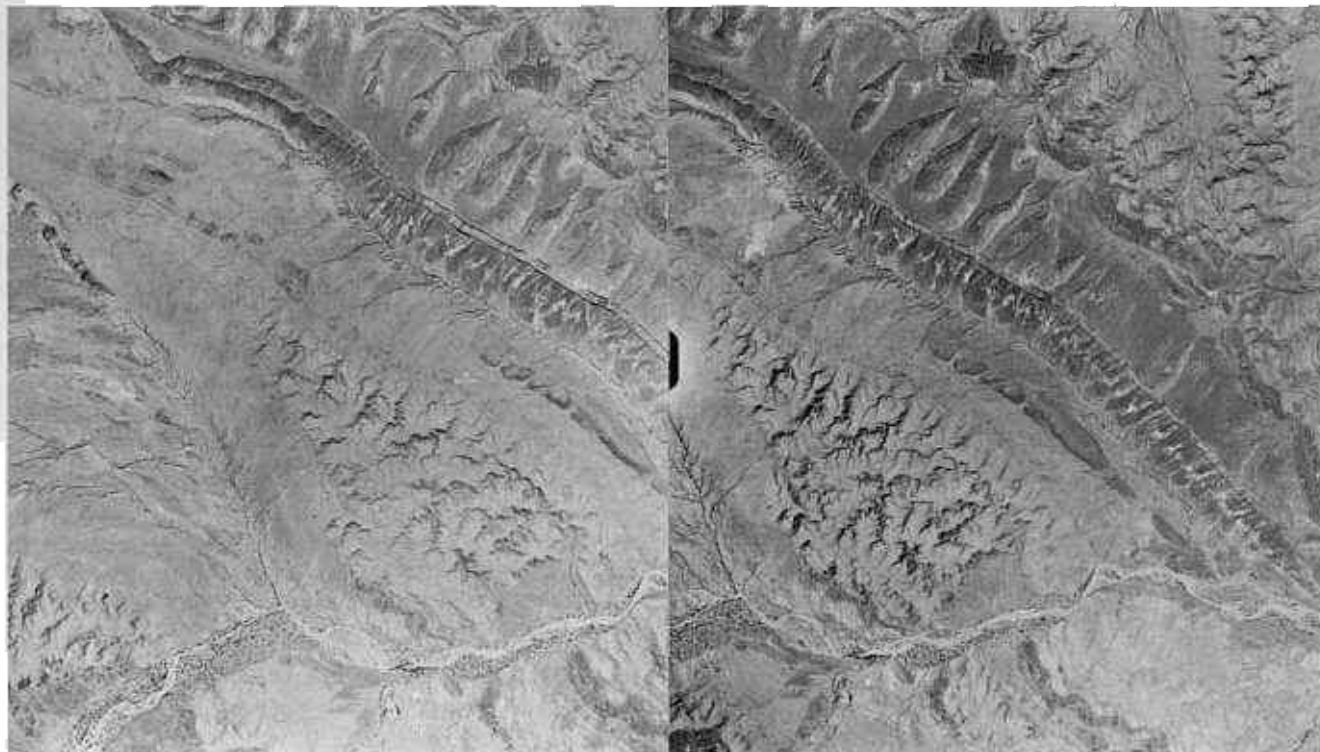


Fig. 4.9 Anhydrite-rich evaporites in this area of Somalia display many of the peculiar landforms that make them so easy to recognize in semiarid terrains. The dip slope in the overlying limestones shows an excellent example of changing

outcrop shapes controlled by varying valley-bottom slope superimposed on uniformly dipping strata. Scale: unknown. Source: unknown.

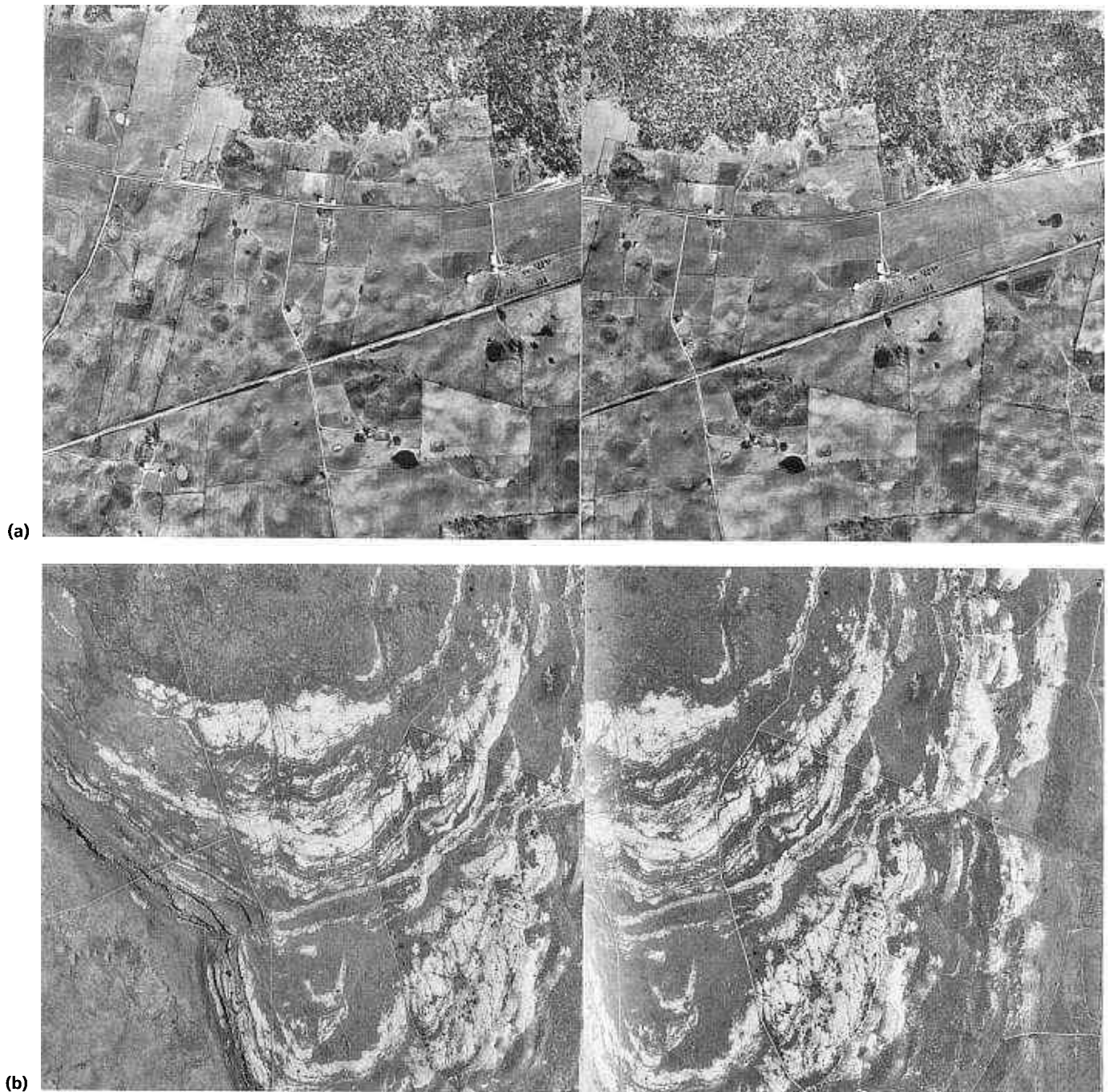


Fig. 4.10 Prolonged chemical weathering of carbonates in humid climates leaves a thin residual veneer of clay-rich soil, which is often quite fertile, as in Kentucky (a). However, the extremely good internal drainage may preclude all but the largest surface streams. The often low-lying surface is generally pock-marked by many depressions over sink-holes. In recently glaciated areas, such as northern England (b), soil development has not been possible over limestones

despite very humid conditions. Limestones have resisted glaciation better than associated sandstones and shales, to form prominent crags. Drainage is almost absent and the bare rock surface is rilled with innumerable solution-widened joints. Vegetation occupies only those surfaces veneered with glacial till. Scales (a) 1 : 25 000 and (b) 1 : 10 000. Sources: (a) USGS GS-XS-6 38 + 39 and (b) Metropolitan County of West Yorkshire 42-68, 176 + 177.

clastic rocks may protect the underlying salt to produce weird pillars and canyons. Being crystalline, evaporites in arid climates tend to resist erosion quite well compared with the fine clastic rocks with which they are frequently associated. They generally stand out as positive topographic features. When in the form of diapirs, the domed nature of the intrusions often controls a distinctive radial drainage pattern, sometimes with an annular component related to the associated concentric fault structures.

Limestone, and to a lesser extent dolomite, is slightly soluble in rainwater with dissolved carbon dioxide—a weak acid. In humid climates, therefore, topographic features resulting from solution dominate areas underlain by carbonates. As carbonates generally are very strongly cemented, shrink in the course of lithification and deform by brittle processes they are strongly jointed. Dolomites, most of which form by the introduction of magnesium into the carbonate structure during or after lithification of a limestone, are yet more strongly jointed because the calcite to dolomite transformation is accompanied by a volume decrease. Solution is achieved most easily along joints or at their intersections. Once it begins, joints are accentuated thereby increasing the already high permeability of lithified carbonates. Consequently infiltration of rainfall into carbonates is extremely efficient.

Above all else the distinguishing feature of carbonates in humid climates is the comparative absence of surface drainage. In the case of more easily dissolved limestones this feature is accompanied by abundant near-circular depressions where rainfall most easily seeps into the bedrock (Fig. 4.10a). These sinkholes may assume monumental proportions where underground drainage has carved large caverns which have collapsed. Such solution features sometimes control local centripetal drainage patterns. The net effect is an area of confused drainage on a landscape mottled by sinkholes and related features.

The solubility of limestone in areas having undergone warm humid climates for long periods results in their assuming a low-lying aspect. Dolomites under the same conditions are not dissolved so easily and therefore tend to form positive topographic features, but still have sparse surface drainage because of their high permeability. In humid areas that have undergone recent rejuvenation the hard and compact nature of carbonates, together with rapid headward erosion and the lack of surface drainage, leaves carbonates as long-lived uplands often with steep-sided major valleys.

Carbonates strongly resist glacial erosion. This combined with the closing of subterranean fissures by permafrost during subsequent periglacial conditions imparts a peculiar aspect to carbonate terrains in boreal areas. The preclusion of infiltration during periglacial

conditions allows the development of extensive steep-sided valley systems in a glacially dominated topography of rugged carbonate scars (Fig. 4.10b). Improvement of the climate 'unplugs' the subsurface drainage and the periglacial drainage is left as dry valleys.

Carbonates may vary in purity, but generally contain very little clastic material. As a result they are capped by very thin soils in humid climates, allowing their internal jointed structure to show up as distinctive patterns (Fig. 4.10b). The exception is in areas of very protracted and intense chemical weathering, as in a tropical rain forest, where residual material accumulates. As the main clastic components are clay minerals, the resulting soils are poorly drained and dense vegetation may accumulate on a terrain almost completely lacking in large streams. In many cases the soil is washed into sinkholes, where particularly luxuriant vegetation may thrive.

In arid and semiarid terrains, carbonates become less easy to recognize on photographs. They resist erosion, have thin soil cover, are well-drained and display prominent joint systems. It is easy to confuse them with coarse clastic sediments. On large-scale photographs with good resolution they can be discriminated by their lack of cross-bedding, their often thickly bedded nature and the peculiarities of their drainage. In most cases, however, they can be distinguished from coarse clastic rocks only by their photographic tone or colour. Many carbonates are very pure, and being composed of white minerals generally appear light on photographs. There are anomalies though. Some limestones have a high content of hydrocarbons and their dark coloration may be confused with that of basic igneous rocks. Such organic-rich limestones accumulated under reducing conditions and sometimes contain iron sulphide grains. Weathering and formation of limonite in these cases can impart strong reddish and brown stains to outcrops. Whatever the initial difficulties in arid terrains, the building up of general signatures for each of the dominant rock types will eventually produce sufficient differences for the carbonate component, if any, to be outlined.

4.2.2 Igneous rocks

Igneous rocks take on a variety of guises because they have a wide range of compositions and grain sizes and may be in various forms as extrusive or intrusive masses. Much effort in photogeology is assigned to establishing the relationships of intrusive bodies to country rocks, which is dealt with in Section 4.3. The most distinctive features of extrusive igneous rocks are those associated with their accumulation at the surface as constructive landforms (Section 4.5). The emphasis here is on distinguishing igneous rocks from sediments and from one another after they have been eroded.

Extrusive Igneous Rocks

The greatest problem in identifying the products of volcanic activity is that they are frequently interbedded with sediments as parallel-sided layers. The problems are removed if they are still associated with relics of volcanic landforms, such as dissected volcanic cones. Then the lavas and associated pyroclastic rocks form roughly circular features with radial and annular drainage patterns, and the only problem lies in discriminating between lavas and pyroclastic deposits.

The classic cone-shaped volcanic edifice characterizes the viscous and explosive eruption of intermediate to acid magmas. With the exception of some areas of within-plate alkaline activity, magmatism of this kind is restricted to long, narrow zones near to destructive plate margins. As well as short-travelled viscous lavas, island-arc and Andean-type volcanism produces abundant pyroclastic materials of several kinds. The coarsest agglomerates travel only very short distances from vents. On photographs they show little difference from coarse clastic sediments, but may be recognized by their restricted outcrops and rapid changes in thickness. Finer grained tuffs, irrespective of their composition and grain size, travel long distances to settle as layered beds of fairly uniform thickness, either subaerially or in bodies of standing water. Because of this they are almost impossible to distinguish from clastic sediments. Their grain size and degree of later cementation determines their response to erosion, on which they can be subdivided in exactly the same ways as clastic sediments. Identification as products of volcanism relies almost exclusively on context—whether or not they are associated with any distinctive volcanic feature.

The exception concerns those pyroclastic rocks that formed from incandescent clouds of volcanic debris with sufficient heat to maintain the particles in a near molten state after they came to rest. In such ignimbrites the particles are annealed together to form a glassy rock, which may contain pumice fragments and abundant pores formed by residual gases. They are indeed tough rocks and form positive topographic features. They can be of any thickness up to several hundred metres. Their porosity ensures that they are well-drained with widely spaced channels. The process of their formation—almost instantaneous deposition of thoroughly mixed debris—ensures that they are homogeneous and lack internal bedding. However, the resolution of photographs determines whether this feature can be discerned. In a dissected terrain ignimbrites therefore can be confused with sandstones and limestones.

Lavas, unless they are undissected and show distinctive surface features (Section 4.5), are difficult to distinguish from sediments with which they may be interbedded. Being crystalline rocks composed of hard

silicate minerals, they usually are resistant to erosion and form positive topographic features. Unless highly jointed, they are impermeable compared with sandstones and carbonates and so develop closely spaced drainage textures. Highly jointed lava flows have excellent internal drainage and may have few surface streams. Joint systems are produced by shrinkage during cooling and tend to be closely spaced with a polygonal pattern, forming columns perpendicular to the flow bases. At the scales normally seen in photographs these patterns are not visible. This structural and compositional homogeneity ensures that dendritic patterns of minor drainage channels occur on their eroded surfaces, in contrast to the angular patterns of jointed resistant sediments.

The principal division of lavas in the field is based on colour, itself determined by the varying proportions of coloured ferromagnesian minerals in the rock. In the fresh state basic lavas are dark, intermediate compositions show as intermediate tones and acid lavas are always light toned. Except under conditions of extreme chemical weathering, the constituent silicates of lavas do not break down *in situ* to release the strong colouring agents of iron oxides and hydroxides. In arid and semiarid climates lavas therefore take on monotonous shades of grey. There is one important exception to this generalization. Some lavas may be altered soon after eruption by percolating hydrothermal fluids. Depending on original composition, the products may be strongly coloured by secondary minerals, including iron oxides and hydroxides. Such altered igneous rocks are of considerable importance in economic geology, because the hydrothermal fluids may have redistributed valuable metals to form ore deposits in association with the altered zones.

Under humid conditions the response of lavas to chemical weathering depends on the proportions of unstable minerals such as ferromagnesian minerals and feldspars relative to stable quartz. Acid lavas are clearly more resistant than basic, and the thickest residual soils form on basic lavas. The minerals quartz and feldspar, which dominate acid and intermediate compositions, contain relatively little in the way of plant nutrient, except in the case of potassium-rich alkali feldspar. Such lavas, except for alkaline varieties, form poor soils and are characterized by thin vegetation. Exposed soils are rich in quartz and clay minerals formed by feldspar breakdown and are pale coloured. Highly alkaline lavas, because of their abundance of potassium and sometimes phosphorus, often support anomalously luxuriant vegetation. More basic lavas break down to good iron-rich soils and support dense vegetation and intensive agriculture. Where soils are exposed they take on a distinctive brown or red hue reflecting high iron contents.

As well as their colour and resistance to chemical weathering, the composition of lavas determines their viscosity and the distance they can flow under gravity. Acid and intermediate lavas are very viscous, flow relatively short distances and as a result display rapid changes in thickness away from volcanic vents. Where thickness can be determined on a photograph, the most acid and viscous lavas often crop out in the form of dissected domes.

The least viscous lavas are those of basaltic composition. They are commonly erupted from huge shield volcanoes or from long fissures, depending on the local tectonic controls over magmatism. They can flow for long distances over the gentlest of slopes to form thick piles of uniformly thick flows. For obvious reasons such lava piles are known as flood basalts. They form nearly level plains before dissection. Once erosion is underway the flood basalts, being resistant, produce plateaux, hence their other common name plateau basalts. The time lapse between individual flows often allows soils and other sediments to develop between flows. The alternating contrasts in resistance then produces a stepped topography, each step representing an individual flow. This feature is equally common in alternating sequences of resistant and weak sediments, however. The pervasive columnar jointing in flood basalts makes them particularly well-drained so that they have coarse dendritic drainage textures. Where incised by streams, they maintain steep slopes and often cliffs owing to the vertical weaknesses induced by columnar jointing. The close joint spacing also ensures that cliffs maintain a very irregular serrated appearance, in contrast to the blocky aspect of sandstone and limestone cliffs.

Because they are so like sediments when eroded, illustrations of dissected pyroclastic rocks are of little use here. Their more distinctive forms in volcanic landforms are shown in Section 4.5. Figure 4.11 shows some of the distinctive features of dissected lava flows in arid and humid climates.

Intrusive Igneous Rocks

The photographic tone or colour of intrusive rocks varies in the same way as that of lavas, and so too does their support of soils and vegetation. The main distinguishing features of intrusions derive from their shape, their internal structures and resulting photographic textures, and their structural relationships with the rocks that they intrude. These relationships are described in detail in Section 4.3. Here it is sufficient to state that they must in some way indicate that bodies of igneous rock post-date those surrounding them to be proven intrusions.

The shape of intrusions when they outcrop at the surface is as diverse as the means by which they were

laced. They may be narrow, long strips resulting in the planation of vertical parallel-sided sheets or dykes. Diapiric plutons take on a variety of forms, ranging from elliptical bodies to irregular masses with rising stocks, bosses and apophyses, depending on their three-dimensional shape and the depth of erosion. Various kinds of ring intrusions take on annular forms, as may laccoliths and lopoliths. There are numerous other minor variants. The most difficult to identify as intrusions are sills, because their emplacement parallel to sedimentary bedding allows them to be used with lava flows or even resistant sediments. However, their form, to describe them fully from photographs the boundaries of intrusions with their country rocks must be continuously traceable. More often this is not possible. Estimates of outlines from photographs have to be backed up by detailed mapping of the ground, including measurements of the attitudes of boundaries.

Figures 4.5 and 4.12 show a variety of the more common intrusive forms.

As intrusions are emplaced at depth within the crust, where temperatures are elevated by the intrusions themselves and by the geothermal gradient, they cool more slowly than lavas. This enables them to developarser crystalline textures than lavas. Cooling also produces joints through shrinkage. The slower the cooling the more widely spaced are the joints. The orientation of the resulting joint systems is often complex, but is largely controlled by the outer surfaces of the intrusion through which heat flows because of temperature gradients. Dykes frequently illustrate this last factor very well, showing regular joints perpendicular to their margins. These patterns therefore give valuable information on the depth of emplacement, the speed of cooling and the shape of the intrusion.

Grain size, as well as composition of intrusive rocks is an important factor in controlling how they respond to weathering. Different silicates expand and contract to different degrees during heating and cooling. This is an important process during physical weathering. Similarly, they change volume as they are transformed to other minerals during chemical weathering, thereby assisting the disintegration of the rock. Both processes increase in effectiveness as the grain size increases. So a fine grained intrusion will be more resistant to erosion than its coarse-grained counterpart.

The great variety of compositions, forms and grain sizes of intrusions precludes a comprehensive treatment. As an illustration of most of the important points in the extensive description of granitic rocks, among the most common intrusions, is given here.

Granitic rocks contain high proportions of the resistant mineral quartz. They are also homogeneous, owing to the mixing processes involved in the convective uprise

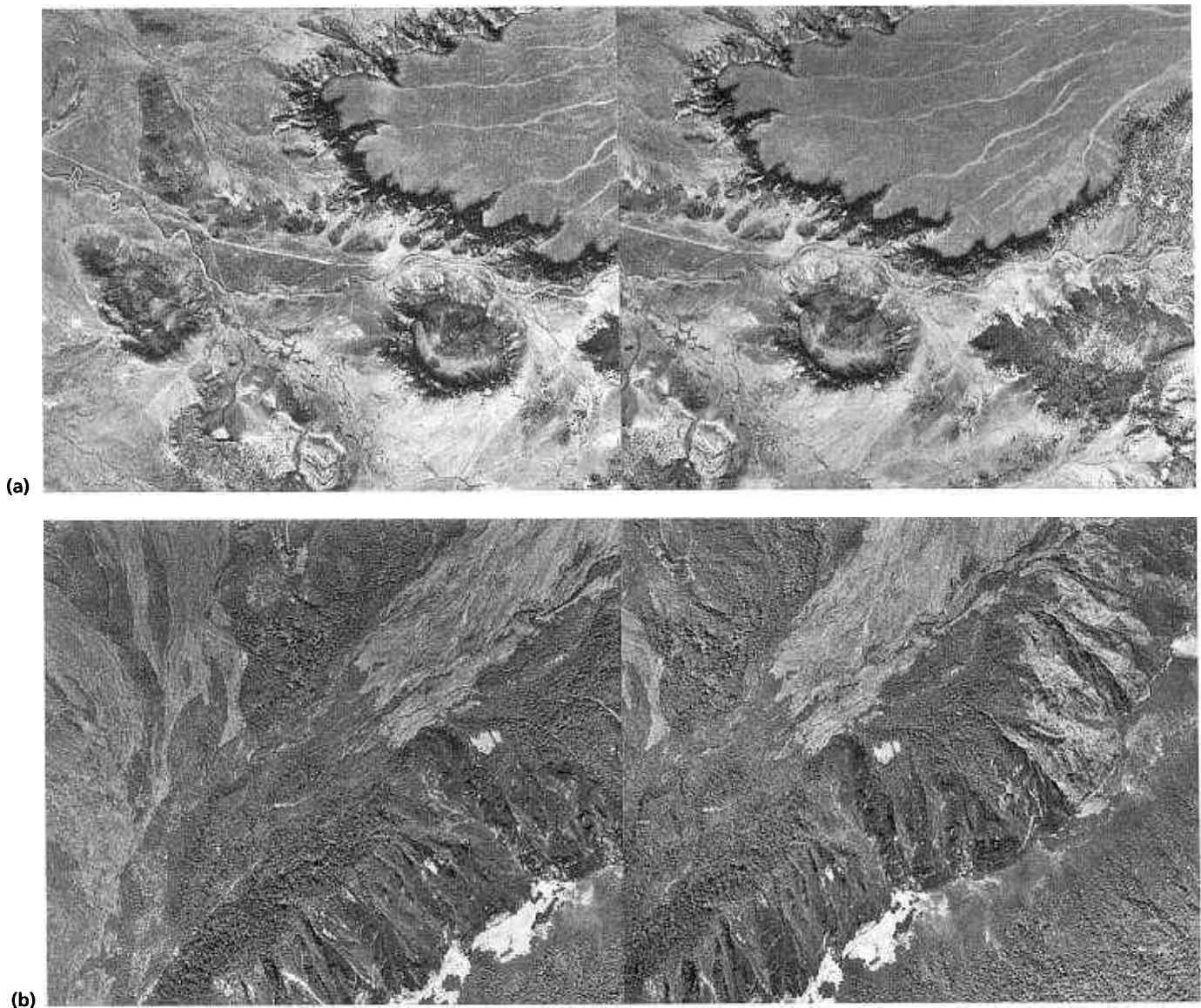


Fig. 4.11 The high resistance of dissected lavas in arid terrains, such as that of Arizona (a), results in plateaux with flat tops and steep bounding cliffs. Isolated outliers often assume the form of spectacular mesas and buttes. Also shown are light-coloured buttes developed on diatremes where the disruption of strata can be seen. In humid tropical areas, such as the Cameroons (b), the high nutrient content of weathered lavas results in luxuriant vegetation, which masks all but the most

obvious primary features of the flows. The central part of the stereomodel shows subdued flow ridges in the pale grey area and possible flow fronts in the dark, jungle areas. Abundant drainage channels together with their association with clear volcanic features suggests that the deeply dissected parts of the area are underlain by lavas too. Scales: (a) 1 : 55 000 and (b) 1 : 40 000. Sources: (a) NCIC Ariz.12 B + C and (b) CSL Series C 45 + 46.

of granitic magmas. As a result granites tend to resist erosion strongly. There is a notable exception. In arid climates, where weathering is dominated by physical processes, such as heating and cooling, the large differences in the coefficients of thermal expansion of quartz and feldspar rapidly cause crumbling of granite outcrops. Under these conditions granites may be expressed by depressions whereas related intrusions of different composition still form positive features (Fig. 4.13).

The homogeneity of granites tends to exert a strong control over drainage networks, which are dominantly dendritic. However, having cooled slowly at depth granitic intrusions have frequently been fissured by large joints which may show as controls of minor drainage. As both form light-coloured outcrops and are strongly jointed, granites and sandstones can be confused. The main distinguishing features are the irregular nature of granitic joint patterns compared with

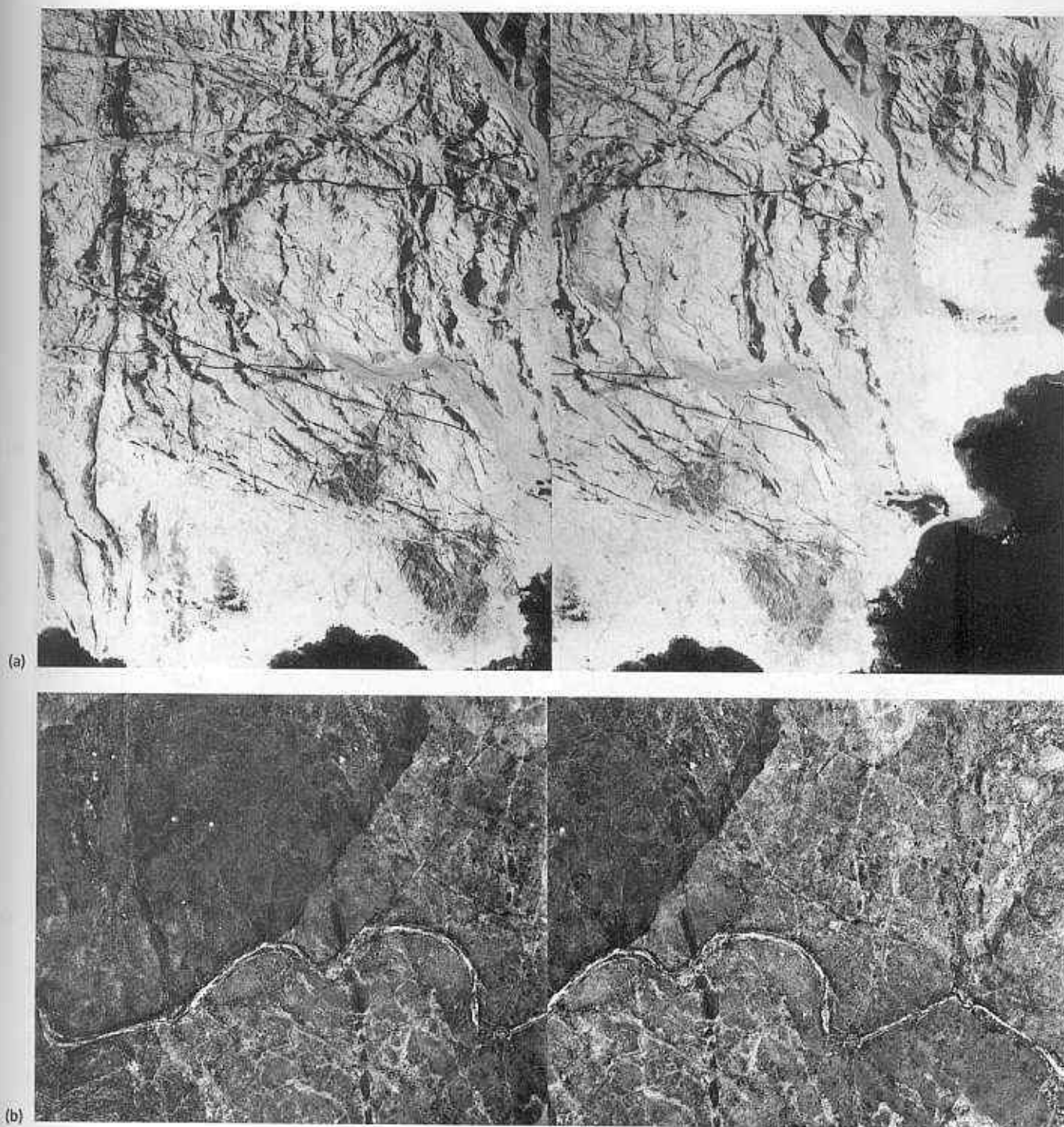


Fig. 4.12 (a) This stereopair of aerial photographs shows an area of light-coloured Precambrian granitic rocks in the Sultanate of Oman. They are cut by several dark, minor basic intrusions. Those with a narrow and nearly straight outcrop are vertical dykes. Intrusive sheets with a low angle of dip are recognized by their sinuous trajectories and by their irregular outcrop widths. The greyish patches in the lower part of the pair are irregular ultrabasic intrusions. (b) The effects of

spheroidal weathering of basic rocks in humid climates and rolling of boulders derived in this way often broadens and masks the outcrops of dykes. In this example from Zimbabwe the dyke is easily recognized where it cuts light Precambrian gneisses, but is only discernible from the dark metavolcanic unit when viewed stereoptically. Scales: (a) 1 : 20 000 and (b) 1 : 40 000. Sources: (a) courtesy of Amoco Inc. and (b) CSL Series C 39 + 40.

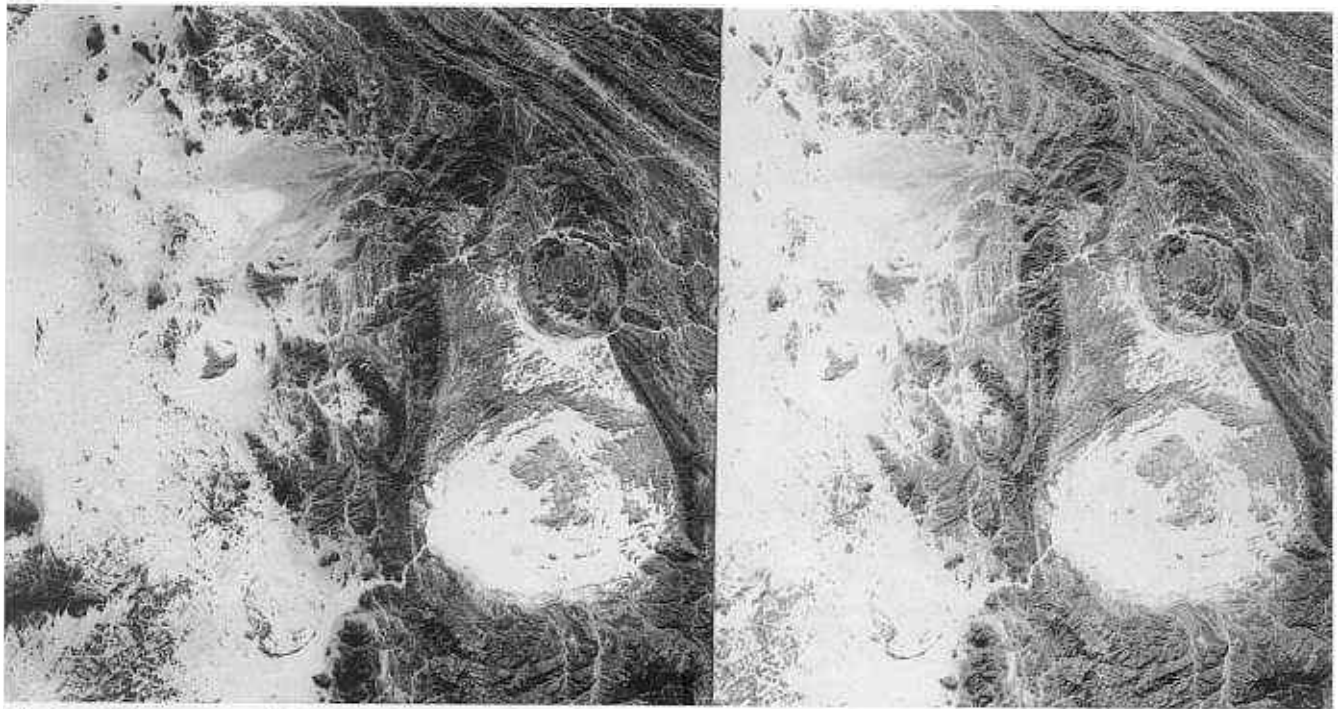


Fig. 4.13 This stereopair of Shuttle Large Format Camera images of part of the Red Sea Hills in the Sudan shows circular granitic diapirs cutting deformed metasedimentary rocks in a late Precambrian shear zone. Some of them have been deeply

eroded to form sand-filled depressions, sometimes with an outer rim of resistant rock, which possibly is gabbroic in composition. The small dark intrusion, clearly post-dating the nearby pale mass is a composite alkaline body. Scale: 1 : 1 million.

rectangular systems in sandstone, and the irregular elevations of granite surfaces compared with those of sandstones, which are controlled by bedding planes (Figs 4.7 and 4.14a).

The crystalline nature of granite, together with its tendency to form poorly bound soils and its hardness means that rainfall runs off efficiently, stripping the surface without forming gullies. Soils may develop in joint-controlled depressions, so that in temperate humid climates bare surfaces criss-crossed with dark lines of thin vegetation are common. The overall effect is a rugged topography of rounded rocky knolls with widely spaced dendritic drainage (Figs 4.5 and 4.14b).

Granites are capable of storing elastic strains produced in them during cooling. They are only released when erosion brings the rock close to the surface, when they are expressed by joint systems that closely parallel pre-existing topography. Further erosion exploits these weaknesses by a process known as onion-skin weathering or exfoliation. In its most extreme form, under humid tropical conditions, this results in the formation of immense domes known as inselbergs, typified by the Sugar Loaf of Rio de Janeiro in Brazil. Inselberg topography is characterized by arcuate drainage patterns between individual domes. Even in areas affected by intense chemical weathering and the development of substantial soils and vegetation upon granites, their distinctive landforms

generally show up (Fig. 4.14c).

Basic and ultrabasic intrusions exhibit many of the characteristics of granites on photographs, particularly in temperate, humid climates. However, bare outcrops are, of course, much darker. Another distinguishing feature of basic intrusions and ophiolitic masses is their common banding (Plate 5.9), which results from the greater tendency of more fluid basaltic magmas to undergo efficient fractional crystallization. The jointing in such rocks is generally more closely spaced than in granites, and in arid climates this is expressed by extremely jagged topography (Fig. 4.15). All the minerals in basic and ultrabasic rocks are prone to decomposition under warm, humid conditions. This produces deep soils with luxuriant vegetation, and a tendency to more subdued topography than that underlain by granites.

4.2.3 Metamorphic rocks

Metamorphism is by definition the growth of new minerals in what were previously sediments and igneous rocks under the influence of elevated temperatures and pressures. Metamorphic rocks therefore may be of virtually any composition. The varied responses to erosion of sediments and igneous rocks by virtue of their composition and grain size, however, is largely masked by the growth of new interlocking minerals

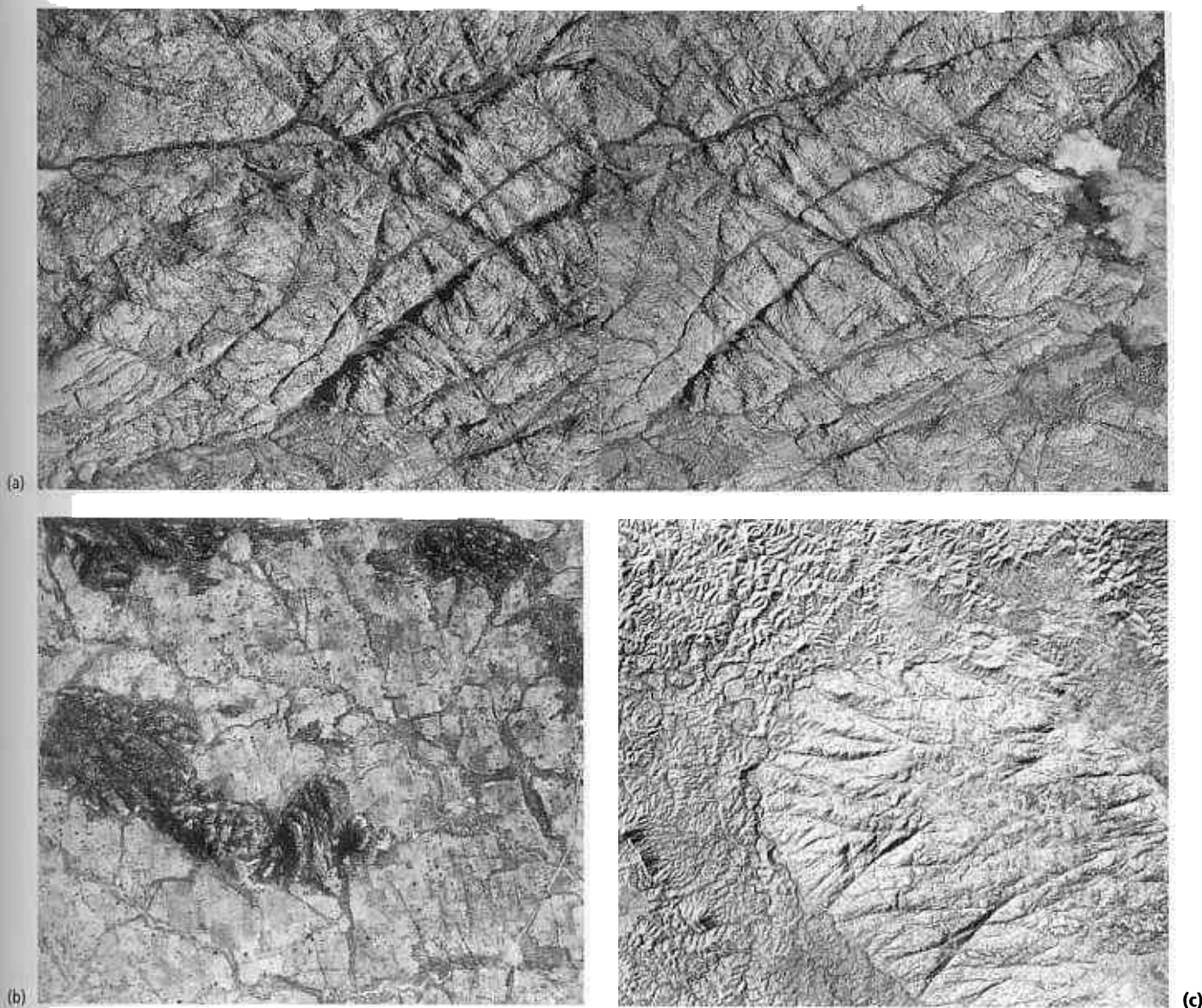


Fig. 4.14. The stereopair in (a) is from an arid part of Wyoming, and shows the typical features of granites eroded under these conditions. The rugged, bare surface is reminiscent of sandstones under similar conditions, but the irregular nature of the more easily eroded joint system and the lack of layering are definitive. Drainage is controlled by the joints, which contain the only vegetation in the area. Images (b) and (c) show the salient features of granitic

rocks eroded under humid conditions in Nigeria and northern Australia, respectively. Image (b) shows a typical inselberg surrounded by a flat pediment of debris derived from the granite. Scales (a) 1 : 37 000, (b) 1 : 30 000 and (c) 1 : 25 000. Sources: (a) USGS GS-CMA-32 88 + 89, (b) Federal Ministry of Works, Nigeria 37-6 68 + 69 and (c) DNDE CAB-151-25125 + 5126.

during metamorphism. Except for moderately soluble marbles, metamorphic rocks are impermeable. They do not contain well-developed joint systems into which rainfall can infiltrate, having cooled slowly during uplift from the deep crust over millions of years. As a result drainage is both finely spaced and dendritic (Fig. 4.16a).

Being composed of strong interlocking crystals, metamorphic rocks are tough and so form positive topographic features. In humid temperate climates, where

they are vegetated, very few distinctive characters show up compared with sedimentary rocks (Fig. 4.16a). In areas of bare rock, in deserts or glaciated terrains, however, they have very distinctive attributes. Most metamorphic rocks contain compositional banding on a variety of scales, representing original layering of sedimentary or igneous origin. Alternatively it may have resulted from the redistribution of the rocks' constituent elements during metamorphism, by the processes of pressure solution during deformation or partial melting at deep

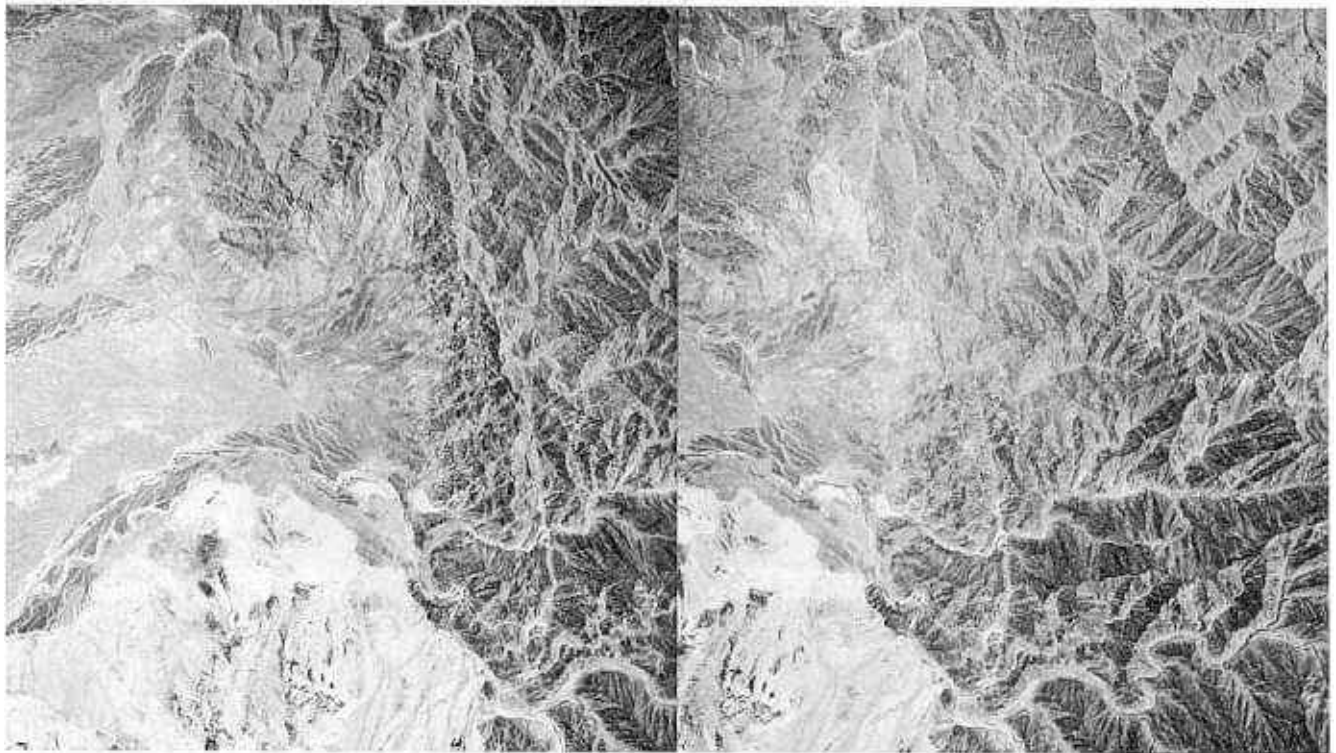


Fig. 4.15 The mountains of north-east Oman are dominated by exotic masses of oceanic lithosphere—mainly gabbros and peridotites—thrust over sediments of the continental margin. In this stereopair the Oman ophiolite shows as a dark jagged

area with a complex and fine-textured drainage system. Its boundary with the block-strewn outcrop of white carbonates is a folded thrust. Scale: 1 : 40 000. Source: I.G. Gass, Open University.

crustal levels. This can impart striking striped textures to bare metamorphic rock surfaces. The influence of the intense deformation which almost inevitably accompanies all but contact metamorphism modifies these patterns into even more distinctive shapes. Repeated folding and shearing results in complex structures, which may be etched out by weathering or glacial action (Fig. 4.16b,c).

It is only under conditions of strong chemical weathering that the compositional variations in metamorphic rocks manifest themselves in different resistances to erosion. As in the case of igneous rocks, the response is controlled mainly by the relative proportions of quartz and other stable silicates, compared with the amounts of unstable, usually ferromagnesian silicates. Siliceous metamorphic rocks are virtually inert and form strong positive features. In a sequence of metasediments these are not necessarily metamorphosed sandstones, but can equally be former siltstones or mudstones in which the original clay minerals have become inert aluminosilicates. Layers with a high iron and magnesium content tend to weather easily. These contrasting responses to chemical weathering are important, because they are the only means whereby the distinctive structural features of metamorphic rocks show up in densely vegetated rain forest (Fig. 4.16b).

Under the intense directed stresses involved in the folding of metamorphic rocks newly forming platy and rod-like minerals tend to orientate themselves, and small masses of rock slide by one another by shear deformation. Both processes result in cleavage, usually parallel to the axial surfaces of the folds. Cleavage forms on scales from millimetres to metres, and only shows clearly as strong parallel linear textures on photographs with the highest resolution. It frequently pervades large areas of metamorphic terrain, however, and may control minor topographic features and small-scale drainage, thereby masking the compositional banding and folds. This is particularly common in glaciated terrains, where the most subtle weaknesses are etched out by the scouring action of ice. Because glaciation produces finely spaced grooves it is easy to confuse these with cleavage (Fig. 4.5).

4.3 Stratigraphic relationships

The keys to successful interpretation of geological history from image interpretation are no more than a few rules used in stratigraphy and the interpretation of geological maps. As in rock identification, observations should be marked on images in some way and details of the criteria used noted down.

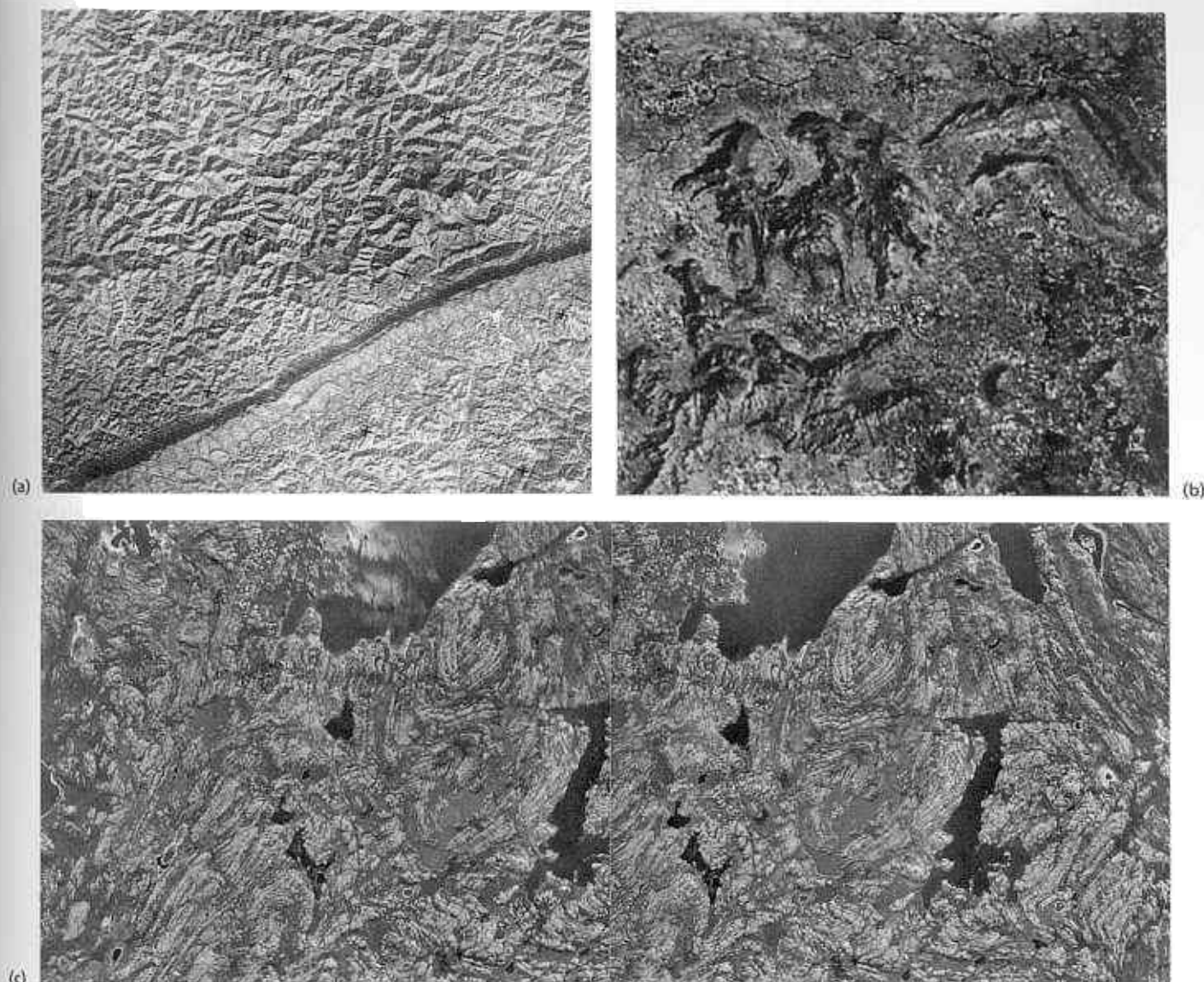


Fig. 4.16 The dissected metamorphic rocks of the Cumberland Plateau in eastern Kentucky show an irregular pattern of dendritic drainage on a Landsat RBV image (a). Very few clues to the underlying structure are revealed at this scale, except in the younger sedimentary sequence running SW-NE across the image, which shows V-shaped outcrops related to the direction of dip. A Landsat MSS band 7 image (b) of the Archaean metamorphic terrain of South India benefits from the selective weathering of rocks of different composition.

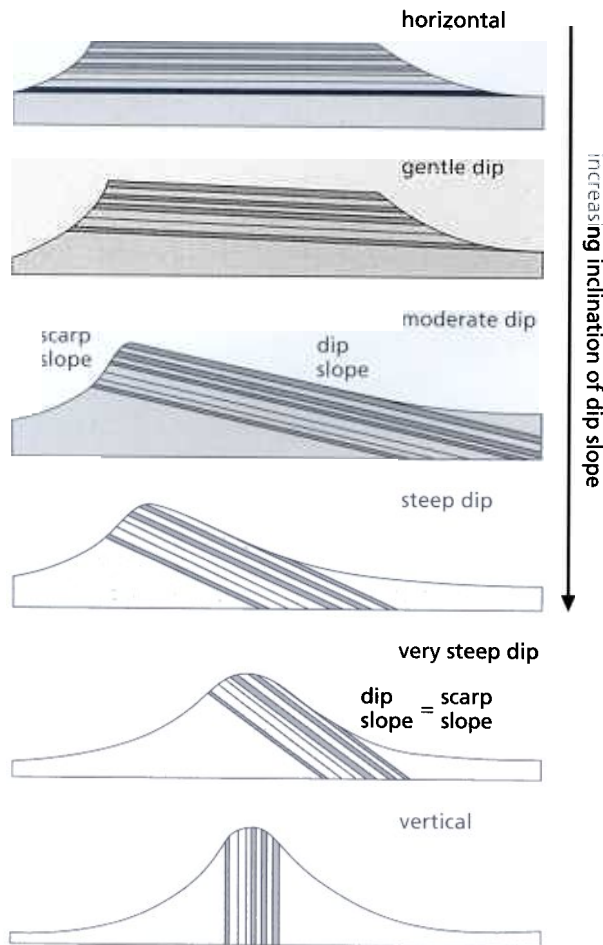
Several of the resistant ridges define complex patterns of folding that typify metamorphic rocks. This diagnostic feature is far more obvious on aerial photographs, particularly of glaciated metamorphic terrains, such as the Canadian Shield (c). The stereopair shows an excellent example of fold interference patterns. Scales: (a) 1 : 500 000, (b) 1 : 250 000 and (c) 1 : 50 000. Sources: (a) courtesy of J.P. Ford, Jet Propulsion Laboratory, (b) author and (c) EMRC A13781 19 + 20, copyright HM the Queen in Right of Canada (1953).

4.3.1 Dip and strike

In the field the angle and direction of dip on flat surfaces, such as bedding, joints, faults and intrusive contacts, can be measured with fair accuracy using a compass and clinometer. On photographs these measurements depend on such surfaces having some detectable topographic expression. In layered sequences, differences in

resistance to erosion may result in the formation of cuestas of one kind or another (Section 4.1.1). They provide the best opportunities for estimates of dip.

Along a cuesta maintained by a resistant layer the slope that has the same direction as the dip is usually a consequent or resequent slope (Section 4.1.1). One with a direction that is opposed to dip is an obsequent slope. Where dip is less than about 30° the obsequent slope

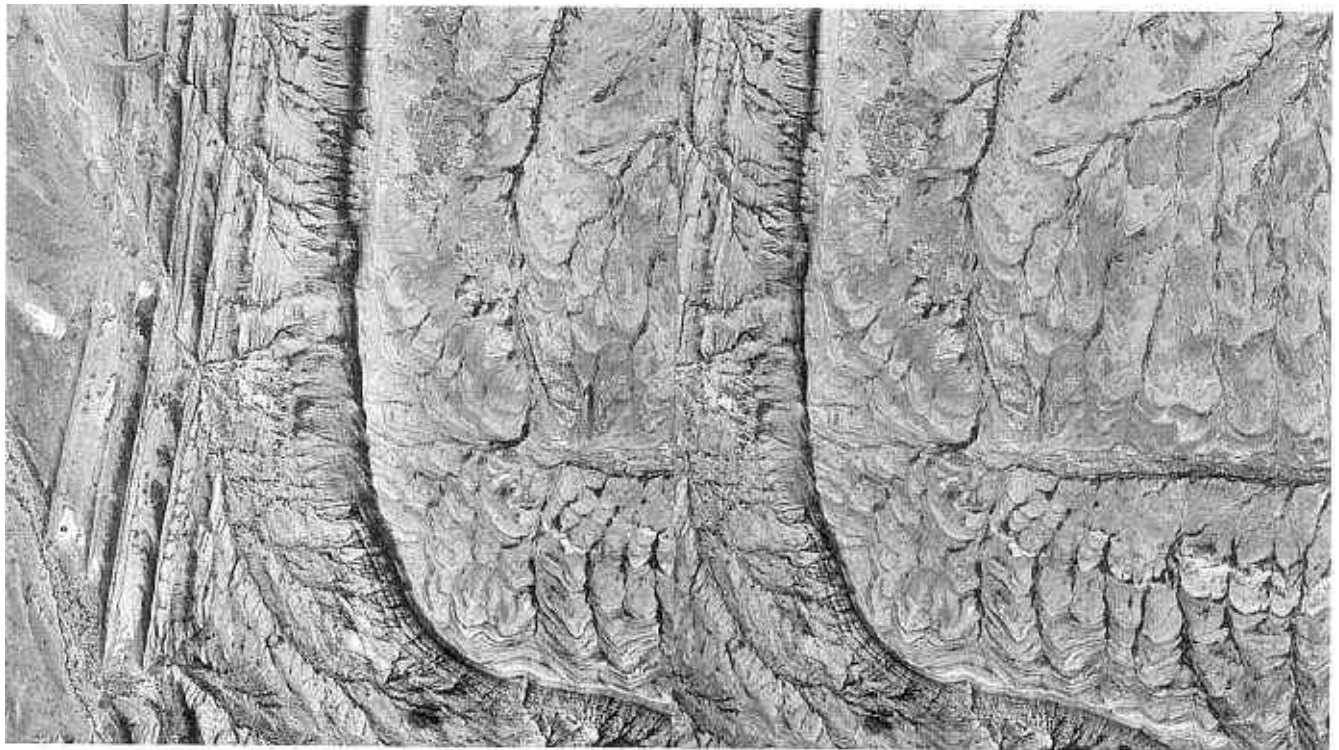


is usually steeper than the resequent slope (Fig. 4.17). This can give a good indication of the direction of dip. However, there are exceptions. Asymmetric cuestas can develop along faults (Section 4.4.1), but this is usually obvious. If the layers immediately above the resistant unit have been stripped from the resequent slope, then the resulting dip slope can provide a very good estimate of the average dip of the resistant layer. Making a dip estimate from photographs depends on being able to view the slope in a stereomodel.

As dip angle increases there is less chance that higher layers in a sequence are completely stripped from resequent slopes. As a result the cuesta becomes more symmetrical (Fig. 4.17). With steep dips the accumulation of talus on the obsequent slope and the increased opportunity for higher layers to slide under gravity down the resequent slope means that the asymmetry of a cuesta can be the reverse of that associated with shallow dips. Although angles of dip can be measured quite accurately from stereomodels (Appendix A), it is a tedious process.

Fig. 4.17 (left) The relationship between the surface form of cuestas and the dip of underlying layered sequences.

Fig. 4.18 (below) This stereopair of an area in Wyoming shows how the outcrop patterns of dipping strata are controlled by both angle of dip and topography. The structure displayed is a monocline with vertical strata to the west and shallow northward dips in the east. Scale: 1 : 32 000. Source: USGS GS-VHI-2 54 + 55.



In most photointerpretation the stereomodel is used to estimate dips into broad ranges— $< 10^\circ$, $10\text{--}25^\circ$, $25\text{--}45^\circ$, $> 45^\circ$ and vertical. Because of vertical exaggeration and radial relief distortion, care has to be taken to apply rough corrections to the apparent dips observed. Figure 4.18 illustrates dip estimation from cuestas and dip slopes. Appendix A includes details of how dips may be measured more accurately from stereopairs.

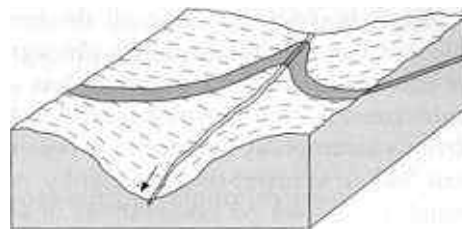
If stereoviewing is not possible, or if clear dip slopes are absent, accurate estimates of dip are not possible. Topography superimposed upon sequences of dipping layers, however, results in the outcrops of layer boundaries taking on shapes familiar to anyone who has studied a geological map. In valleys the boundaries assume V-shapes, the Vs pointing in the direction of dip. On interfluvies the boundaries are more arcuate, with concavities indicating dip direction (Fig. 4.19). In extreme cases, the landforms so produced take on the appearance of flatirons. Using the law of Vs is essential in rapidly assessing the variation in direction of dip over large areas. Where folds are not expressed by distinct shapes it is often the only means of detecting them and distinguishing antiforms and synforms. Figure 4.18 gives an example of this useful technique. Occasionally it is possible to use the law of Vs on single satellite images.

Where dips are steep, or in areas of subdued relief, valleys do not result in V-shaped outcrops. Traces of layers and their boundaries provide only an indication of strike. This is not reliable in the case of shallow dips in subdued terrain, because the slightest slope has a marked effect on the trend of outcrops. Where resolution is poor or where suitable boundaries do not cross suitable valleys, strike determination is about the best that the interpreter can expect to achieve. Provided the elevation differences are small compared with the scale of the photograph, or that visible units form ridges or valleys of roughly constant elevation such estimates of strike are fairly accurate. This generally is the case for satellite photographs. However, in deeply dissected mountains even this aid to interpretation is notoriously unreliable.

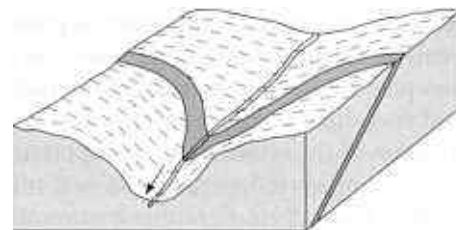
Another method of assessing the direction of dip, where valley-Vs are absent, is by examining drainage patterns. As Figs 4.1 and 4.17 indicate, where dips are less than about 45° consequent and resequent streams tend to be longer than obsequent streams. Provided such streams are present, mapping them out can reveal a crude direction of dip.

Dip and strike estimations are not restricted to boundaries in layered rocks. Faults, the margins of igneous intrusions and unconformities are all surfaces for which geometry is important to geologists. Unconformities frequently form cuestas and their dips can be estimated easily. Igneous intrusions, such as batholiths, generally form uniform positive features relative to the rocks that

(a) Upstream dip



(b) Steep downstream dip



(c) Downstream dip less than stream gradient

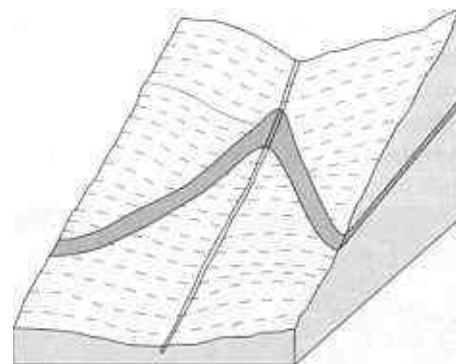


Fig. 4.19 These block diagrams show how the interplay between topographic slope and the dip of strata control the shapes of outcrops in valleys and on spurs. Generally, V-shaped outcrops in valleys point in the direction of dip, but if topographic slope is locally greater than a dip in the same direction, the relationship is reversed (Fig. 4.9). The apical angles of the Vs become more obtuse as the dip becomes steeper, until no topographic displacement of outcrop occurs with vertical layers.

they intrude. Cuestas are rarely involved. However, the contact may display Vs in valleys and reveal its direction of dip, and perhaps clues to the dip angle. As plutonic igneous contacts are often highly irregular, information of this kind may be misleading. Provided it is not vertical, the hade of faults may be revealed by V-shapes in valleys.

4.3.2 Superposition

If the dip direction of any geologically important surface is definable on a photograph then it is a simple matter to work out the vertical sequence of rocks cropping out on either side of it. In a layered sequence the rock cropping out in the dip direction is above the surface, that cropping out in the opposite direction is below. In

most cases the rock above the surface is younger and the rock below is older. By using all the boundaries between layers that are expressed on a photograph it is therefore possible to build up a more or less complete stratigraphic column of the most obvious rock units. There is one obvious assumption, that there has been no major recumbent folding to invert the stratigraphy. Assessing this possibility is based on observations of structural complexity and on context. Field observations, which are vital to fill in the detailed stratigraphy, will in most cases show whether tectonic inversion is present from way-up indicators. Figures 4.2, 4.9, 4.16(a) and 4.18 give some examples of simple stratigraphy which can be established once dip direction is known.

Identification of the relative age of important or prominent units in a layered sequence is not only useful in establishing a rough stratigraphic framework, it also is a useful aid to recognizing the presence of folds and their sense of closure. The relative stratigraphic position of units juxtaposed across faults allows the sense of movement to be established, and in some cases the amount of throw too.

For faults with hade less than 90° exactly the same principle is used to discover the hanging wall and foot-

wall. In terrains dominated by thrust tectonics, such as the Rockies of western North America and the European Caledonides, and by nappes as in the Alps, recognition of thrust surfaces and their dip directions is the key to establishing a vertical sequence of tectonic units (see Section 4.4).

4.3.3 Unconformities

Unconformities are among the most important stratigraphic features because they represent periods of time separating episodes of deformation, uplift and erosion from the re-establishment of deposition. They allow the geological history of an area to be subdivided into convenient episodes. By definition an unconformity separates older, more complex sequences from stratigraphically younger rocks with simpler structure that lie above the unconformity. This usually is revealed by the overstepping of the unconformity surface from one level in the underlying sequence to others. Overstep shows on photographs as an irregular and discontinuous line, marking the unconformity, which truncates traces of bedding and other structures that lie topographically below it (Fig. 4.20).

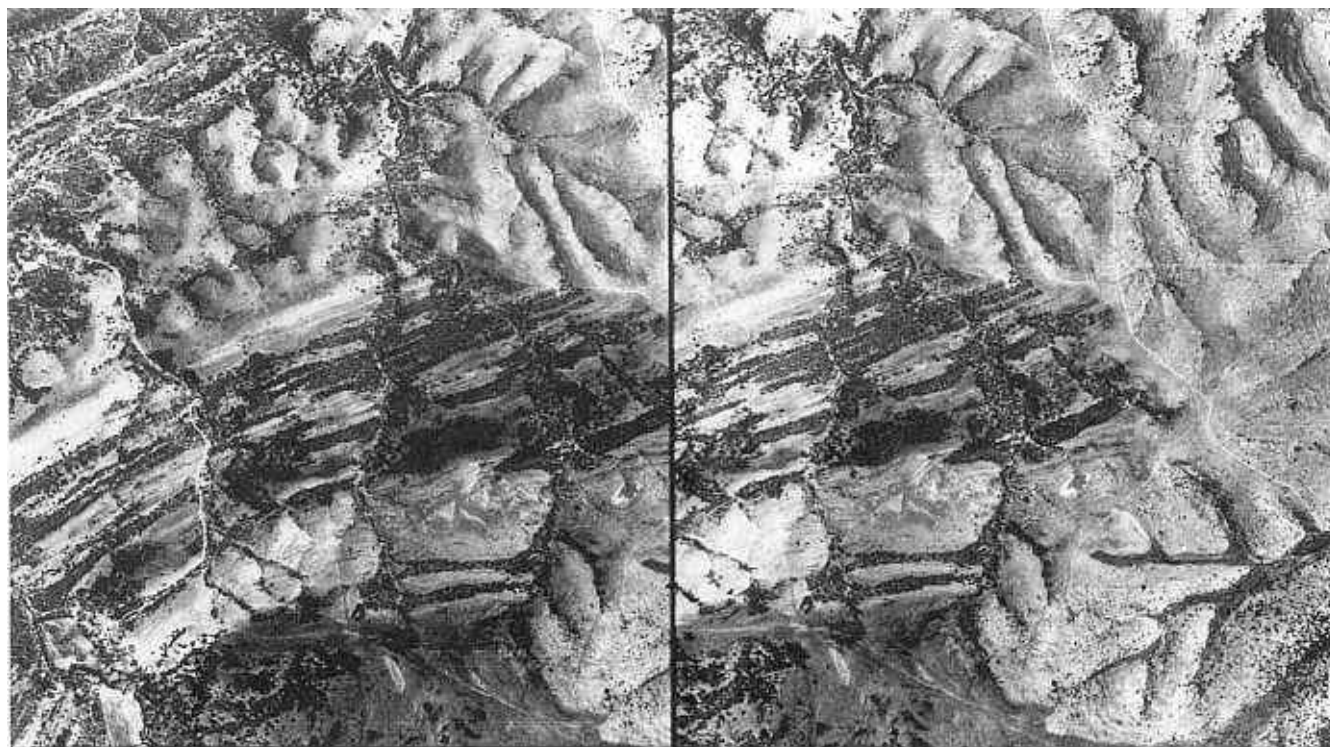


Fig. 4.20 On this stereopair of aerial photographs of an area in Oklahoma can be seen two distinct units. One is characterized by a dendritic drainage pattern and is sparsely vegetated. Faint traces of bedding parallel to contours suggests that it is horizontal. The other appears banded, with several almost straight, wooded ridges, which appear to be controlled by steep dips to the north-north-west. The boundary between

the two clearly truncates the ridges, and the horizontal unit lies unconformably upon the steeply dipping strata. The wide spacing of drainage in the younger unit suggests that it is a massive, coarse clastic rock. The older unit comprises shales and limestones. Scale: 1 : 22 000. Courtesy of the Committee on Aerial Photographs, University of Illinois.



Fig. 4.21 A mosaic of several of the Landsat MSS band 5 scenes surrounding the area in Plate 4.2 reveals the complete regional setting of the late Precambrian Cuddapah Basin. In the west it is a shallow undeformed basin, but in the east has clearly been affected by strong deformation. Dark, forested ridge-forming units display complex fold patterns, and in fact the whole basin has been inverted and overthrust by Archaean basement from the east. Scale: 1 : 2.5 million. Source: author.

Unconformities are usually near-planar boundaries, marking the presence of erosion surfaces that have developed to completion. Where erosion is interrupted early in the cycle of denudation by sudden increase in the local sediment load, a rugged landscape may be buried and preserved as a very irregular unconformity. Only examination of a stereomodel will allow such buried landscapes to be evaluated.

Although unconformities permit the ages and intensities of tilting, folding and faulting to be related to the overlying rocks, their most revealing relationships are to igneous intrusions. As igneous rocks can be dated by radiometric means, their age relative to the unconformity, as well as to the rocks and structures divided by it, can place absolute time constraints on local geological history (Plate 4.2).

Unconformities can be tilted and folded, in which case their attitude can be estimated using the same methods as those applying to dipping layers (Section 4.3.1). Because they are major time boundaries separating rocks of contrasted structure, unconformities can often be traced over huge areas. Satellite photographs with their super-synoptic cover are then very useful. Because elevation differences are effectively 'ironed out' at the small scales

involved, the regional shape of the outcrop of an unconformity can be used to estimate quite subtle features relating to original basin extent, and the effects of regional warping and faulting younger than the unconformity (Fig. 4.21).

4.3.4 Cross-cutting relationships

As mentioned in Section 4.3.3, igneous rocks generally are the only units in an area that can provide reliable radiometric ages for calibration of the stratigraphy. Intrusions give the minimum age of the rocks that they cut and a maximum age for those which sit unconformably upon them. Similarly, they allow the time relationships of structural features that they transect, or which disturb them, to be assessed. Sedimentary sequences are built up over millions of years, with perhaps very limited evidence of age such as fossils or radiometrically dated units. They provide only crude age limits for tectonic activity. Intrusions however, are emplaced over geologically short periods of time and provide a more sensitive means of working out tectonic evolution.

Apart from sills and other intrusions with margins concordant to older layering, all intrusions have margins that cut across older structures, as shown in Figs 4.5, 4.13 and Plate 4.2. In the case of massive intrusions it is rare to find evidence of their later deformation. Narrow dykes are particularly sensitive to distortion, however, and provide near-ideal markers for structural investigations.

4.4 Structural relationships

Folds, faults and thrusts, and their relationship to features such as unconformities and intrusions are used to work out a qualitative tectonic history for an area. They also aid quantitative estimates to be made of the displacements involved and the disposition of principal stresses. Using structures as indicators of the mechanics of deformation is largely beyond the scope of this book. Only a few important comments are made, full details being available in standard textbooks on structural geology. In this section the emphasis is on detecting and describing major structures.

4.4.1 Faults, lineaments and arcuate features

Faults are surfaces across which blocks of crust have moved relative to one another. Identifying a feature on a photograph as a fault therefore means establishing that displacement has indeed taken place across it. Except for thrusts, most faults dip steeply and crop out as more or less linear features that are relatively unaffected by topography. There are, however, many ways in which obvious linear features may be formed. They may be

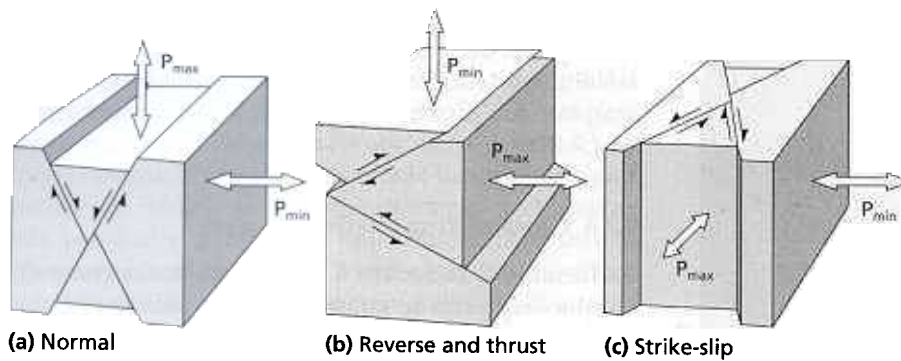


Fig. 4.22 These block diagrams show the relationships between the three main classes of fault and the required arrangements of the maximum and minimum principal stress axes. Theoretically, for each arrangement of principal stresses, faults should develop with two orientations. The acute angle between them is bisected by the maximum principal stress direction, and the two sets form a conjugate pair.

faults, joints, dykes, steep to vertical strata and cultural features, such as roads and boundaries between areas of different agricultural use. Certain defects in some images produce artificial lines too, known as artefacts (Chapter 5 and Appendix B). It also may be possible to link various elements of a scene by lines, such as isolated boundaries between surfaces of different tone, meanders in a number of streams, notches in ridges, truncated spurs and features relating to other geological structures. These may be purely fortuitous connections, but could have some genetic relationship to subtle or deeply buried faults.

With the increasing availability of small-scale, super-synoptic satellite images of the Earth there has been a tendency among some geologists to cover them with real, inferred and entirely imaginary lines. In some cases these extend across entire continents. Repeated analysis by the same person or by different interpreters often results in different arrangements of many of the lines. The near obsessive nature of this simplistic approach has provoked a measure of scepticism and cynicism towards the whole of remote sensing by many scientists. There can be no doubt that linear features of geological importance do occur, and that they are visible only on small-scale photographs because they form broad, subtle and very long connections. However, all potential lines need to be viewed with caution and subdivided into categories with different degrees of confidence (Section 2.6).

A note on nomenclature is necessary here, to avoid adding to the already burgeoning confusion surrounding lines on images. Linear is an adjective describing the line-like character of an object or array of objects. It is often misused as a noun designating clear, short lines on an image, in contrast to a lineament which is correctly applied to long, often subtle linear arrangements of various topographic, tonal, geological and even geophysical and geochemical features. Linear feature is used here to describe relatively short (less than 5% of the extent of the image being used) lines and regular arrangements of surface phenomena. It is used together with the term lineament in an entirely nongenetic sense. When either can be recognized or confidently inferred as a geological entity, then it is given its correct, geological name.

Faults manifest the brittle response of the crust to stress. They are surfaces along which the crust has failed and which allow the stresses involved to be dissipated by crustal extension or shortening. The attitude adopted by faults is governed approximately by simple mechanical laws relating the orientation of the maximum, minimum and intermediate stress axes acting in the crust, the strength of rocks and the resulting surfaces of potential failure. Figure 4.22 illustrates these principles and shows the main classes of faults that develop. At the instant of failure one of two possible fault surfaces begins to break first and strain is concentrated and propagated across it. Thus, although theoretically faults should develop equally, parallel to two conjugate directions, in most cases one direction dominates. The conjugate set of faults to this is either suppressed or plays a relatively minor role in crustal deformation.

Heterogeneities in the crust and in the stresses involved during a period of crustal deformation mean that all the types of fault shown in Fig. 4.22 may occur in an area. If they do, then estimates of their attitudes and the directions of movement along them allow the overall tectonics of the area, and their relationship to major crustal forces to be assessed (Fig. 4.23). An important parameter to assess from such composite fault systems is the overall

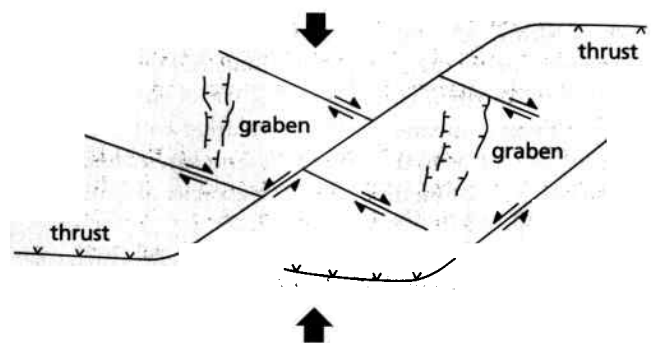


Fig. 4.23 Because the Earth's crust is constrained in two dimensions, one major episode of deformation, such as that characterizing the Tibetan Plateau today, can produce all three classes of faults. This cartoon shows how they might form relative to one another.

sense of movement of blocks of the crust that the faults bound. This is often difficult for normal and thrust faults. For strike-slip movements, the relative displacements visible may allow both dextral and sinistral movements to be recognized. Dextral means that the movement sense of a block on one side of the fault relative to an observer on the other side is to the right. Sinistral faults displace to the left. The orientation and sense of fault movement provides the basic information from which the disposition of the regional principal stresses can be worked out. This in turn gives a context within which other, more subtle structural features can be sought.

The majority of faults that are easily visible on images have steep to vertical hade, as explained later. However, many change hade with depth in the crust to form what are known as listric surfaces, from the Greek for 'spoon-like'. As a result their surface expression may change with the depth of erosion. No fault continues indefinitely through the crust. It may die out, the strain being taken up along other faults. It may split or develop numerous minor splay faults along its length. Such splay faults result from the reorientation of principal stress axes in the close proximity of major faults. In turn, large splay faults have a similar but less widely distributed effect, and develop minor splays of their own. The result is a complex pattern of curved faults and tectonic joints of different sizes.

The photointerpretation of faults is often more reliable than their detection in the field. Because they are usually weaknesses, fault surfaces are rarely seen at outcrop. They often control low-lying features, from which views into the distance are difficult. Changes in vegetation and surface texture related to faults are difficult to see when the geologist is close to them. The synoptic view of photographs enables widely separated pieces of evidence to be linked as sharp and semicontinuous linear features or lineaments. The vertical exaggeration associated with stereoscopic viewing accentuates even the most subdued topographic features and reveals clearly any vertical displacements associated with faulting. Despite the scepticism about 'linear geoart', image interpretation is the most powerful means of detecting crustal fractures and analysing their tectonic and economic significance.

As with all geological boundaries, the dip on fault surfaces governs how they are affected by landforms. The lower the angle the more a surface tends to follow topographic contours and adopt complex outcrop patterns. Thrusts are controlled strongly by topography, and their expression is often identical to that of bedding surfaces. On occasion the floor of a thrust, which may be coated by tough mylonite, stands out as a prominent dip slope (Fig. 4.24c). Sometimes the tones, patterns and textures of rocks forming the roof of a thrust clearly indicate that they are older than the rocks forming the

floor (Fig. 4.24c). The roof may be metamorphosed basement and the floor composed of sediments with simple structure. Distinctive lithologies in the roof rocks, such as red beds, may also indicate an inverted sequence and the suspicion of thrusting. In the main, however, thrusts are notoriously elusive in photogeology; because they have low angles of dip their outcrop often parallels that of the layers that they have displaced. Often they can be inferred only from the context of other observations. One of these is the identification of local truncation of compositional layering in the hanging wall or footwall (Fig. 4.25). These are known as strike cut-offs, and are more easily recognized on images covering large areas than on the ground, because the angle between thrust and local strike is often very small.

The steeper the angle of a fault the less it conforms to topography. Steep faults are conspicuous on photographs because of their near independence of valleys, ridges and uplands. The steeper they are the more they resemble regular lines. Because of intense fracturing along faults they often form local weaknesses, become eroded out and form linear depressions (Figs 4.24, 4.30 and Plate 4.3). They are often followed by streams. The almost irresistible tendency of streams to meander means that any straight segments immediately raise the suspicion of fault control. There are relatively rare cases of faults that have been cemented by percolating fluids or followed by igneous dykes, in which case anomalous straight ridges can form.

The high permeability of fault zones means that they can be saturated by water. They provide easy access for root systems, and when forming depressions protect plants from wind and desiccation. Lines of vegetation encouraged by these conditions are also potential indicators of faults, but joints provide exactly the same conditions and other evidence is needed for confirmation. Where faults have thrown permeable rocks against impermeable, spring lines may form. In the case of limestone thrown against impermeable rocks, local drainage may disappear into lines of sinkholes marking the fault. Where a fault juxtaposes rocks of grossly different resistance to erosion, the fault can be expressed as a line of cliffs, waterfalls or abrupt changes in slope.

Whatever linear features can be discerned on a photograph, they can be identified as faults only if other geologically significant features in the scene are displaced or truncated. Even truncation is a dubious criterion, because geological structures can terminate abruptly for a host of reasons. Normal and reverse faults displace the crust vertically. In terrains dominated by horizontal or gently dipping strata the displacement shows up only by changes in elevation of easily recognized beds or boundaries across the fault. Where dips are greater than about 5°, vertical displacements also produce apparent lateral

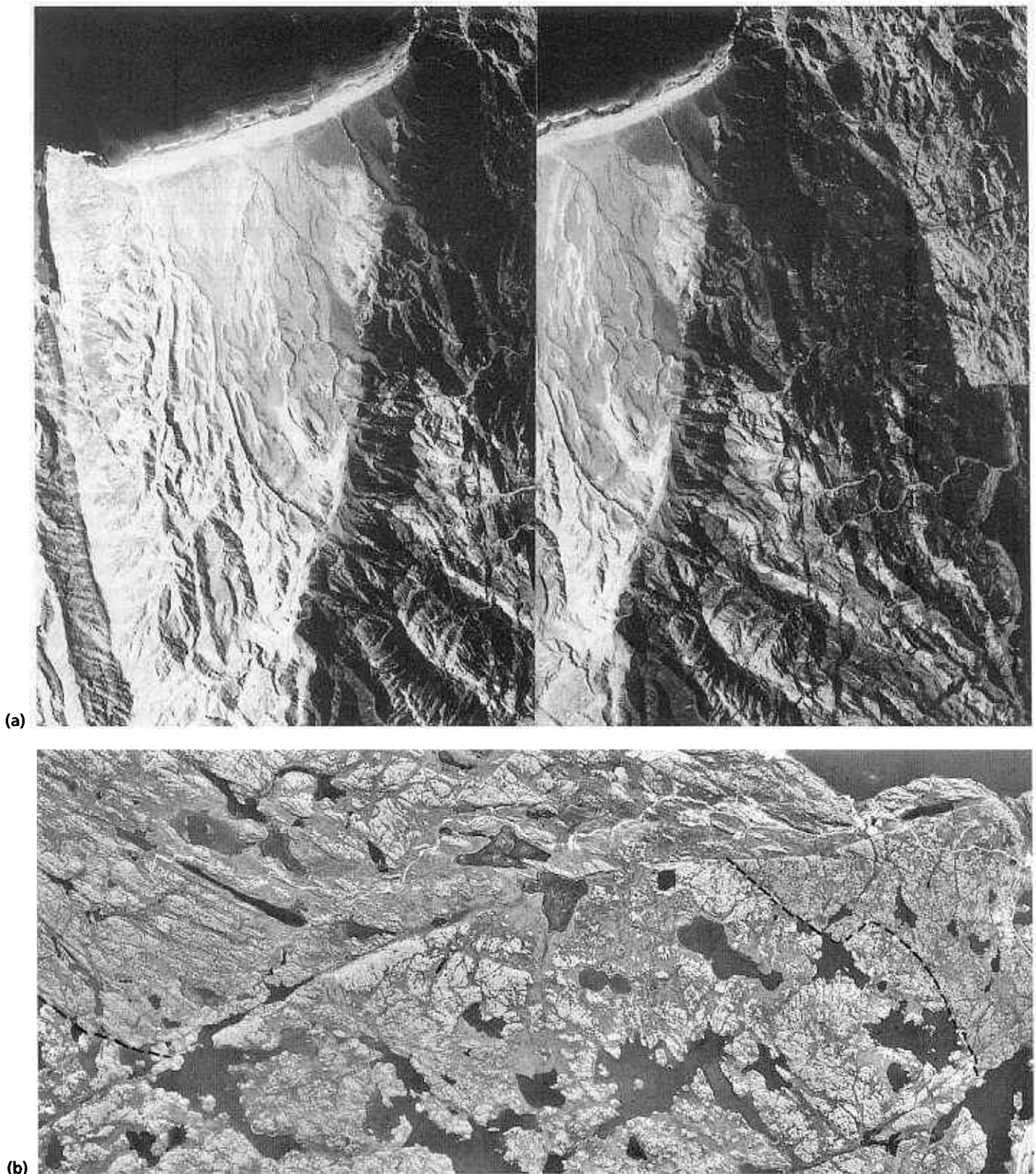


Fig. 4.24 (a) The sharp boundary separating dark from light terrains is a very large normal fault in the Sultanate of Oman. The light unit consists of Tertiary limestones, which were deposited upon the dark, late Cretaceous ophiolite complex. The ophiolite here is composed of layered gabbros in which compositional banding is clearly visible. It has been thrust over the pale-grey Precambrian granites seen at top right.

The thrust itself has been displaced by a small strike-slip fault on the right. (b) The sharp linear feature running diagonally across this aerial photograph of part of the Canadian Shield is a strike-slip fault. The displacement across it can be judged from the relative disposition of the vertical contact between granite (light grey) and metavolcanic rocks (darker grey) marked by the dashed line. It is about 5 km.

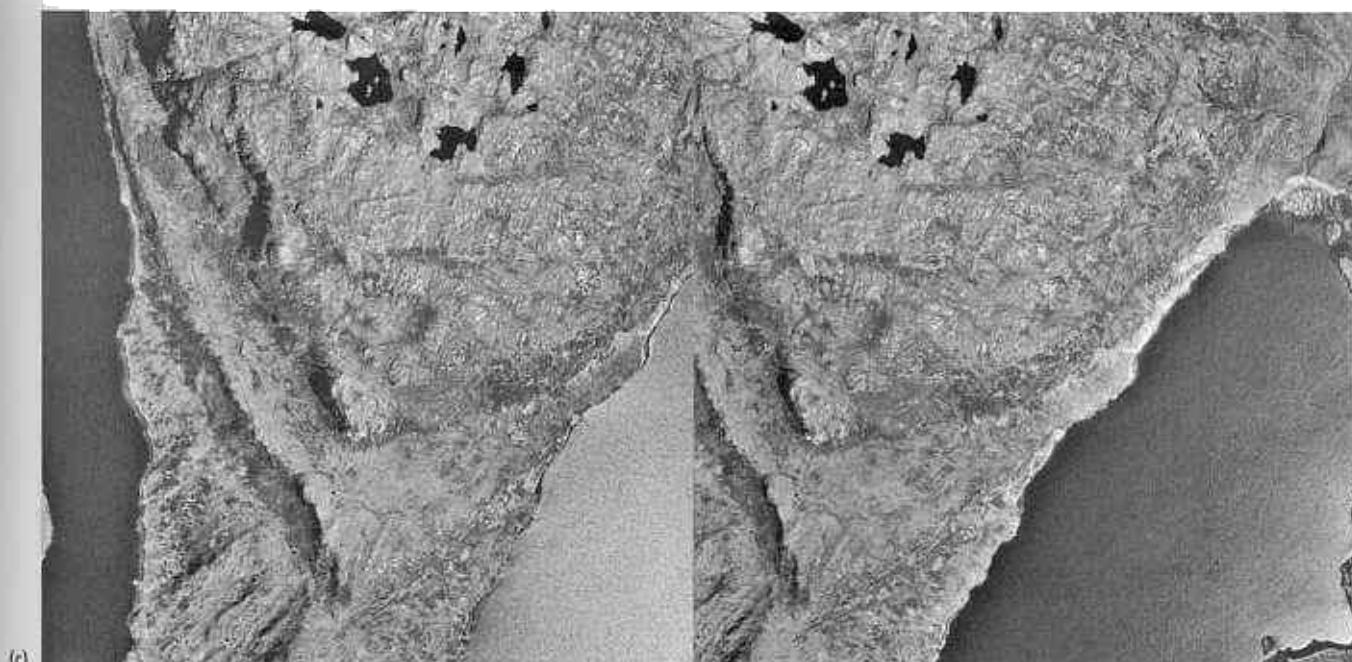


Fig. 4.24 (c) This stereopair shows a spectacular thrust fault in the Caledonides of north-west Scotland, which drove Archaean gneisses over Cambro-Ordovician sediments. The thrust plane is marked by the pronounced bench on the peninsula, which is coated with mylonite. The typical

knobbly topography of the heavily glaciated gneisses is quite obvious in the upper part of the pair. Scales: (a) 1 : 20 000, (b) 1 : 50 000 and (c) 1 : 20 000. Sources: (a) courtesy of Amoco Inc., (b) EMRC A13621 177, copyright HM the Queen in Right of Canada, and (c) author.



displacements in outcrops. Features marking resistant units are truncated by the fault and reappear across it at another position, depending on the direction of dip and the sense of movement of the fault. Knowing the directions of dip of the unit and of the fault surface—perhaps from V-shaped outcrops—allows the sense of faulting to be worked out. Strike-slip faults produce true lateral displacements. However, the resulting truncation and displacement of dipping resistant units can have exactly the same appearance as those associated with normal and reverse faults. The only reliable criterion is the displacement of vertical surfaces, such as other faults, vertical strata, igneous dykes and the axial surfaces of upright folds.

As well as revealing the present tectonics of an area, active faults present major environmental hazards. Locating them and identifying their movement senses is doubly important to the geologist. World-wide monitoring of seismic events has provided such a wealth of data on the location of earthquake epicentres that all

Fig. 4.25 (left) On this Landsat TM band 5 image of an area in Eritrea the presence of a fault almost parallel to compositional banding, probably a thrust, is indicated by a marked linear feature to either side of which the strike of smaller linear features corresponding to layering is truncated. These are ramps in the hanging wall and footwall of the thrust. Scale 1 : 100 000. Source: author.

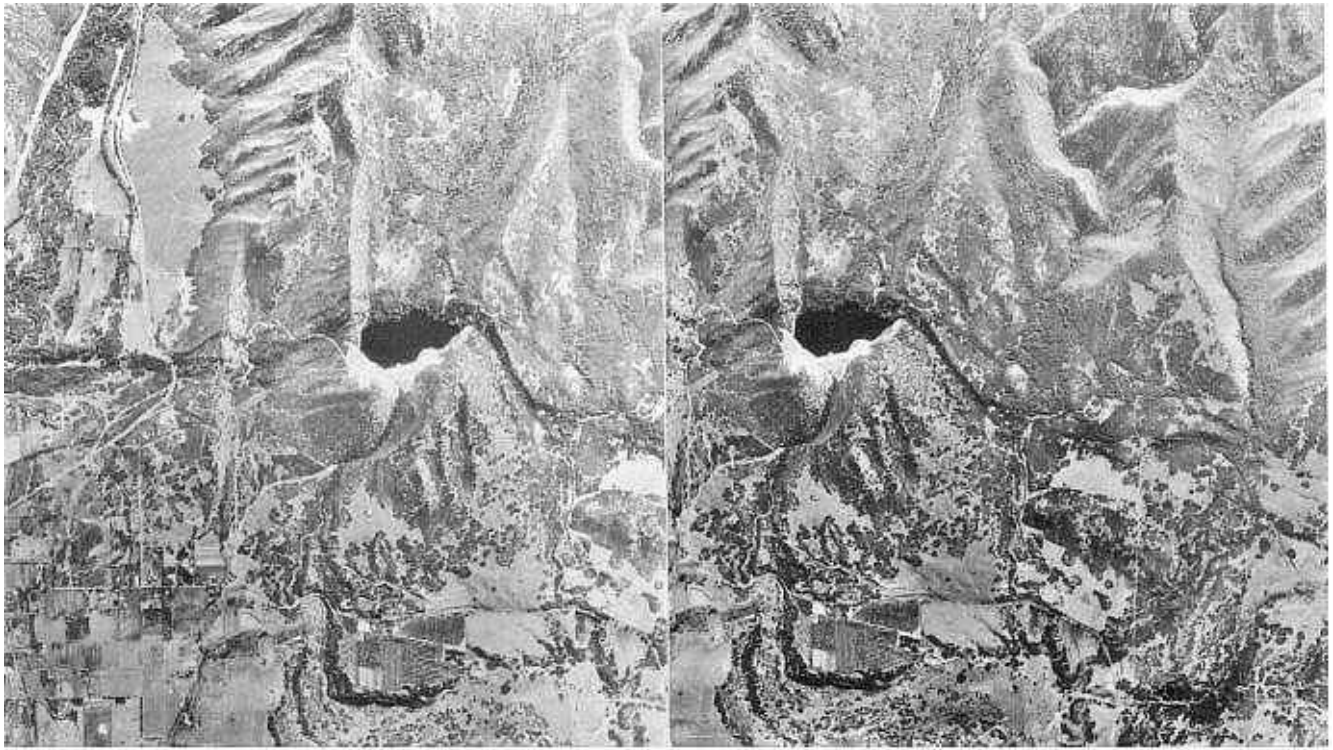


Fig. 4.26 Running across the centre of this stereopair of an area in Utah are several connected scarps. They are developed within Pleistocene moraines, and resulted from normal

displacements along an active fault. The body of water at the centre is a typical sag pond caused by fault-induced tilting. Scale: 1 : 28 000. Source: NCIC Utah 2 B + C.

areas of major seismic activity are now known. Seeking active faults in them therefore is aided by known context. The world-wide increase in potentially hazardous and seismically sensitive nuclear installations, however, means that even the smallest earthquakes pose a serious threat. Currently many areas hitherto considered stable are being re-evaluated for seismic risk. This concentrates on delineating old fault systems that may be reactivated by minor tectonic processes, such as isostatic readjustment after glacial loading.

Large active faults are recognized in the same ways as old ones etched out by erosion. Their activity has extended for thousands, if not millions of years, so that erosion has had time to pick them out. In many cases they are in areas of considerable relief resulting from the rapid uplift and erosion that accompanies intense tectonic activity. They are easily seen in such terrains. An active tectonic zone may also contain faults that have ceased to move. So it is important to distinguish those that pose a threat from those which are probably safe. An active fault may remain dormant for tens and hundreds of years between earthquakes, with stresses slowly being built up before explosive release. There may well be no records of seismic events related to them. Fortunately, faults that have been active in the last few thousand years still preserve unique features which can be seen on photographs.

The most obvious sign of a fault's activity is its disturbance of cultural features, such as roads and straight field boundaries. Equally as clear are displacements of Recent superficial deposits, often with the presence of clear breaks and low bluffs (Fig. 4.26). Small-scale tilting associated with fault movements produces depressions, often filled with water to form sag ponds (Fig. 4.26). Where streams cross active strike-slip faults they are periodically displaced. This produces sudden bends in their courses along the fault line. Such offset streams are often present in all drainage crossing the fault, and define both the line and the sense of most recent movement (Fig. 4.27). Physical disruption by a major earthquake of topographic features formed during periods of quiescence, results in the formation of truncated or faceted spurs (Fig. 4.27). Such displaced ridges can be transported so far along a strike-slip fault that they block the upper ends of valleys, the formative streams of which have been shifted elsewhere. This produces shutter ridges and headless valleys (Fig. 4.27).

The supersynoptic view provided by satellite images is ideal for re-evaluating fault patterns and other reflections of brittle tectonics over large areas. Small, regionally insignificant features are suppressed and very large, subtle and previously unsuspected ones may show up. The smaller the scale the more subtle, more completely connected and larger are the features that can be discerned.

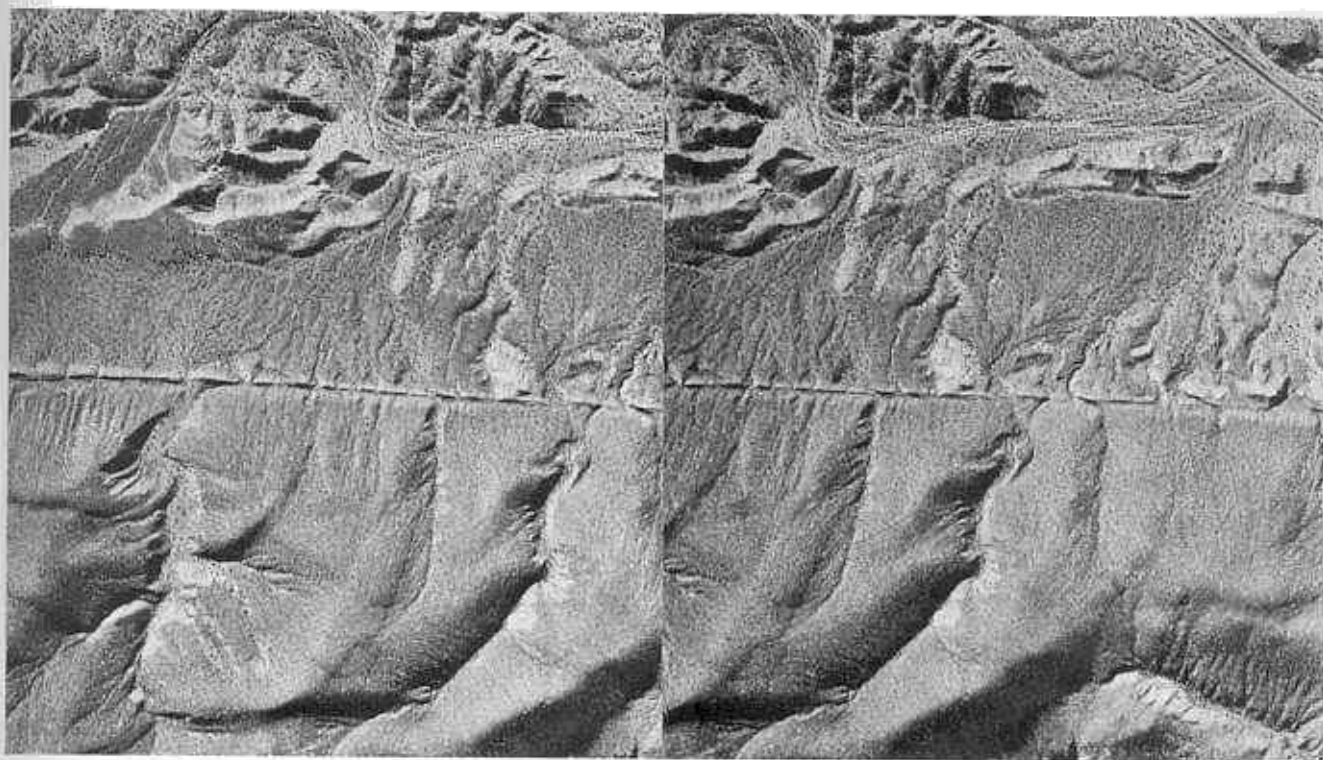


Fig. 4.27 The linear feature running across this stereopair is one of many active faults in California. Along it can be seen good examples of offset streams, headless valleys, a shutter

ridge and faceted spurs. The sense of active movement appears to be sinistral strike-slip. Scale: 1 : 12 000. Source: USGS GS-GF-1 39 + 40 (colour original).

Individual scenes or mosaics of many scenes can be used to highlight the interconnectedness of tectonic features (Fig. 4.28). Bearing in mind the cautionary comments earlier in this section, linear features and lineaments can be mapped very quickly and related to notions of the tectonic evolution of an area. The best results are from areas of high relief and recent activity, where faults are well-expressed and where evidence for movement directions is most clear (Plate 4.3 and Fig 4.29).

As well as their importance in tectonic analysis, faults have economic importance. Being zones of intense fracturing, they afford easy passage to various fluids. Because vertical faults ignore topography, where they cross from areas of high elevation and rainfall to plains with shortages of drinking and irrigation water they may transfer large supplies to areas of need (Fig. 4.30). Fracturing of the crust can also focus the migration of hydrothermal fluids and magmas. Both may encourage mineralization. In the early years of Landsat image interpretation explorationists concentrated on the mapping of linear features and lineaments, and their relation to known mineral districts. This was encouraged by the long-held suspicion that some mines were focused on intersections of deep-seated faults. Further incentive stemmed from the synoptic analysis of regional patterns of mineralization that began to reveal linear arrangements of active mines and mineral prospects. One suggestion was that they



Fig. 4.28 This NOAA-6 AVHRR band 2 image of West Asia shows the relationships between the tectonically linked Oman ophiolite (SW), the SSE-NNW Lut Block—a microcontinent that indented Iran—the E-W folded Makran Range and the N-S folded Kirthar and Suleiman Ranges, which are bounded to the west by a large sinistral strike-slip system. All formed at roughly the same time in the late Cretaceous to early Eocene when the Tethys ocean closed and India impacted Asia. Scale: 1 : 15 million. Source: author.



4.29 This sketch map of Plate 4.3 distinguishes between faults and compositional layering. The N-S faults bound a major pull-apart basin, filled and cored by Pleistocene volcanics. The NE-SW and NW-ESE faults are conjugate strike-slip faults, the dominant type being the former. This pattern has been interpreted as reflecting northward crustal shortening and E-W extension. The crust in fact deforms as a framework of lozenge-shaped blocks bounded by strike-slip faults and which contain the volcanics. Scale: 1 : 1.8 million.

over deep crustal weaknesses. Another was that they recorded the passage of the crust over mantle hot-spots. Theoretically, both mechanisms should be reflected in the tectonic and magmatic features visible at the surface, given a wide enough perspective.

Another regular and sharp type of feature that has attracted considerable attention from explorationists is the circle. The eye is very good at seeing circles. Indeed, it is rather too good, and can pick out circles in nearly any random patterns. Nevertheless, there are many nearly circular and elliptical features produced by crustal processes, and some of them are indeed related genetically to mineralization and hydrocarbon accumulations. The most obvious are salt domes produced by the isostatic uprise of low-density halite and the production of annular fault systems and overlying circular domes (Fig. 4.31a). They are well-known traps for oil and gas. Circular granite intrusions form in a directly analogous way, and when they are peralkaline, have associated tin, niobium and rare-earth mineralization (Fig. 4.13). An important proportion of diamonds are found in kimberlite pipes, the products of cryptoexplosion structures associated with ultramafic magmas rising from the deep



4.30 In Eritrea the western lowlands contain large areas of initially fertile, clay-rich soils yet they record much lower rainfall than the nearby upland areas to the east. Runoff from the uplands is heavily charged with sediment and dams to store it for irrigation would soon become silted. Equally ephemeral streams are so powerful during storms that they would destroy channelled irrigation schemes. On this aerial photograph several marked linear features are visible. These are large crustal fractures trend from the high rainfall uplands to plains with clay-rich soils but low rainfall. Presumably, they could be transferring large volumes of high quality ground water to the area of agricultural potential. The water could be tapped quite easily. Scale: 1 : 700 000. Source: author.

crust. Such pipes too have a nearly perfect circular cross-section. However, kimberlite is very soft and pipe erosion often shatters the country rock so that circular features are suppressed. Within large igneous plutons, of diorite to granodiorite composition, the patterns of convective flow of magma and hydrothermal fluids can be seen in circular cross-sections of flow cells. These may lead to mineralization of the porphyry type. Another example of a real circular feature relates to the erosion of volcanic calderas and their frequently associated annular faults produced in calderas. Many porphyry-type ores are found at different levels in such structures, and may be exposed by erosion (Fig. 4.31b).

A view of the Moon reveals the most dramatic of circular features—the craters associated with the impact of large meteorites. There is no way in which the Earth could have escaped the bombardment responsible for the scarred lunar surface. However, the vast majority

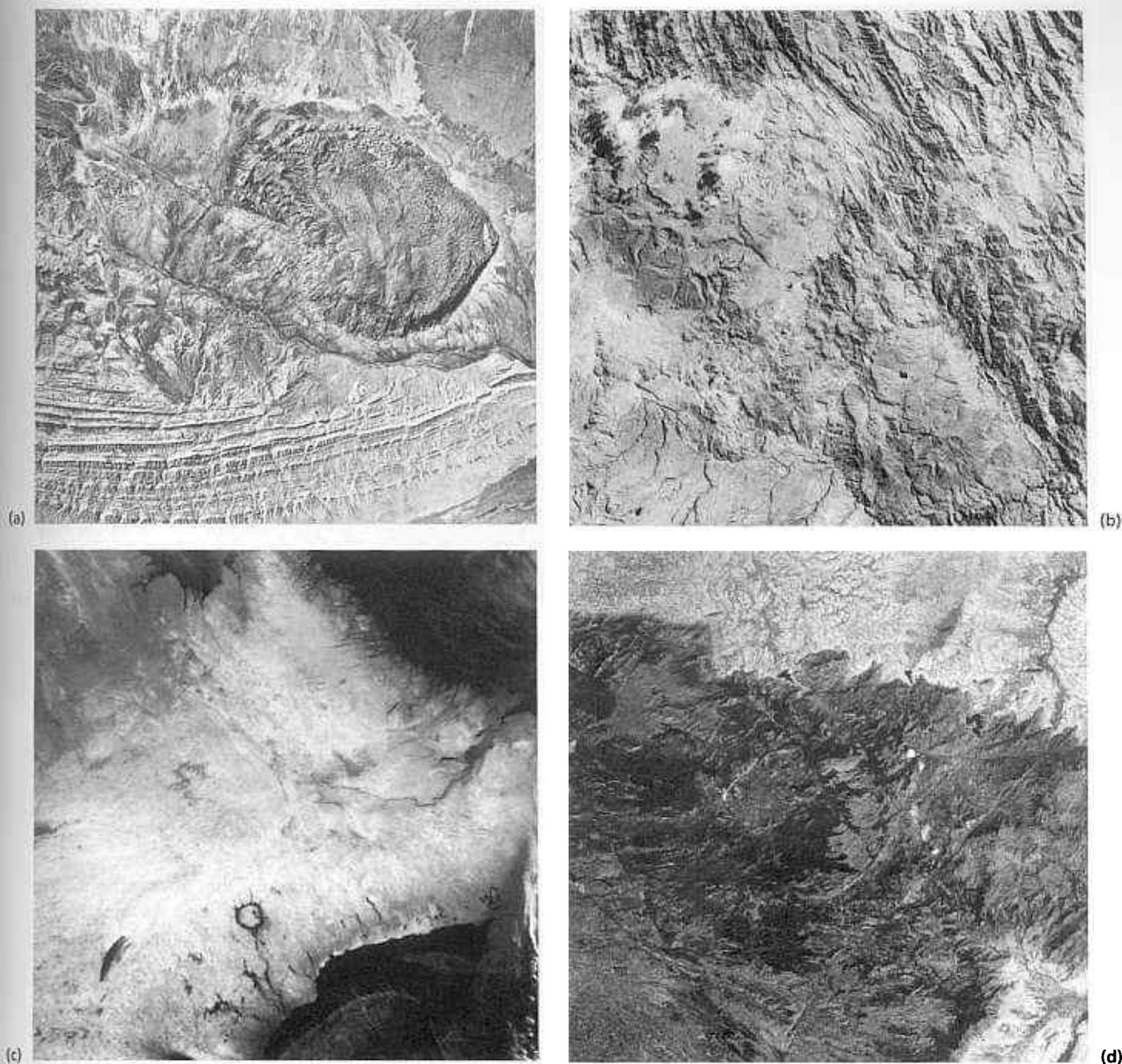


Fig. 4.31 Circular and elliptical features, especially those so large that they are difficult to see on the ground, are easily seen on small-scale images. The smallest are commonly related to salt diapirism like that from Iran shown in (a). On (b), a Landsat MSS band 7 image of part of the Andes, can be seen a circular volcanic edifice some 40 km in diameter. Image (c) shows Labrador and parts of Quebec on a NOAA AVHRR near-infrared image. The obvious circular feature is the Manicouagan meteorite impact, about 70 km across.

of lunar impact structures arose during the period before 3.8 billion years ago. Since then the entire crust of the Earth has evolved and been reworked. All the truly enormous terrestrial impacts—theory suggests the

The circular feature at the centre of (d), a Landsat MSS band 5 image, is located in the granulite-facies gneisses of the Anaimalai Hills in South India. Although it is bounded by a clearly defined boundary, on the ground there is no evidence for what may have produced the structure. Scales: (a) 1 : 90 000, (b) 1 : 1 million, (c) 1 : 14 million and (d) 1 : 1 million. Sources: (a) CSL Series B 11, (b) courtesy of P.W. Francis Open University, (c) National Oceanic and Atmospheric Administration, Washington, and (d) author.

largest to have had a diameter of 6000 km—have been obliterated. Such impact structures do exist where little has happened to the crust for a billion years or more. They are among the easiest features to spot from orbit

(Fig. 4.31c). On an entirely different plane are circular features that are defined by mineral occurrences on a continent-wide scale, and which show up, if at all, only on satellite images encompassing millions of square kilometres. Several authors have interpreted such features as relics of the early bombardment of the Earth, and continually seek others as a regional guide to new mineral districts. This is an entirely illusory quest, because the thermal effect of even the largest impact would last only about 50 million years. Moreover, the mineralization associated with such features affects rocks with age ranges that span up to 2.5 billion years. Whatever these giant features are—and if they are not imaginary—they probably reflect long-lived anomalies in mantle heat production, and their episodic effect on crustal uplift and thermal evolution. Finally, there are circular features that undoubtedly exist, but to which no mechanism can be assigned (Fig. 4.31d).

4.4.2 Shear zones

Near to the surface the crust deforms by brittle processes and by buckling. With increasing depth the influence of higher temperatures and hot fluids decreases the viscosity of rocks, so that they are more prone to deform plastically. Instead of crustal displacement being taken up across sharp fault surfaces, the shear strain is distributed through a large volume of rock, but the net effect is exactly the same. Such deep-seated, ductile fault movements produce shear zones, the forms of which are very different from those of faults. Like faults, shear zones result in thrust, normal and strike-slip displacements of the crust. Their typical form is illustrated in Fig. 4.32.

Deformation in a shear zone involves changes in angular relationships, so that linear features outside the

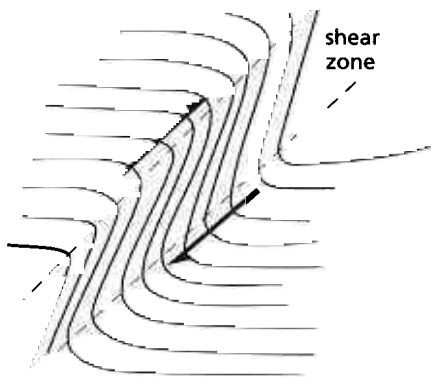


Fig. 4.32 This sketch shows the main elements of a shear zone. Layers are displaced by shearing in the same way as in a fault. However, instead of a sharp break, the displacement is accomplished by rotation, extension and flattening of the layers, so that the shear zone itself comprises highly deformed, banded and often lineated rocks.

shear system are seen to swing into it. This can be seen in Fig. 4.32 to be accompanied by a narrowing in the spacing between the linear features as the shear zone is entered. The reason for this is simply the tectonic environment in which the shearing takes place. The crust is being compressed in the direction of the maximum principal stress and extended at right angles to it (see Fig. 4.22). Some of this crustal deformation is taken up by the displacements along the shear zones, but because rocks are responding plastically, local compressive and extensional strains develop along the shear zone itself. The rock is flattened in the plane of the shearing.

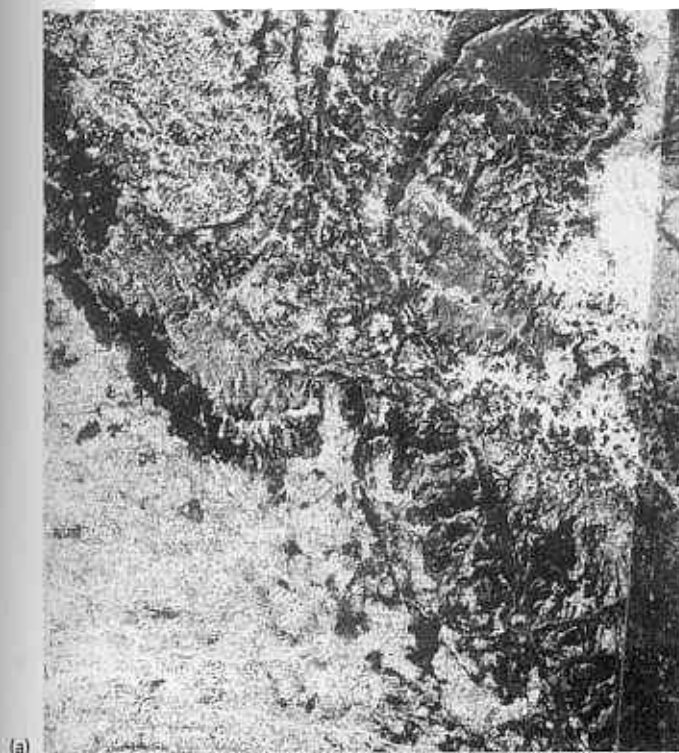
The more strain is taken up by a shear zone, the further displaced are once adjacent points either side of it, and the wider the zone of flattened and stretched rocks becomes. If marker units are present then their displacement is immediately apparent. In most cases this is not possible, but theoretical methods allow the shear strain to be estimated from the width of the belt of flattening. Even if this is not measurable, the sense of strike-slip movement is quite obvious from the way in which linear features are rotated into the shear zone (Figs 4.32 and 4.33).

As well as deforming linear features, shear zones also rotate and flatten folds, often into unique and bizarre forms (Fig. 4.33b). Any relatively narrow belt of finely spaced linear features with flattened folds separating terrains with simpler structure can be suspected of being a shear zone with a strike-slip component. Thrust-like shear zones are not so obvious, but can be detected from their association with complex fold forms (Section 4.4.3).

4.4.3 Folds

Folding of the crust presents the most stimulating and challenging images to the geologist. As well as providing an element of variety for the interpreter, and another means of evaluating strain and stress patterns, folds are of great economic importance. Anticlinal structures form traps for oil and gas, whereas synclines often host artesian water supplies. Locating and analysing folds is an important element in predicting subsurface geology, and in reconstructing the geological history of an area.

Folds may be classified using a great range of geometric properties, for which there are excellent guides in standard texts on structural geology. In photointerpretation most of these criteria are not visible, so here a grossly simplified system is used. It is based on three criteria—the interlimb angle of folds (Fig. 4.34a), the plunge of fold axes (Fig. 4.34b) and the dip of fold axial surfaces (Fig. 4.34c). As the way-up of folded layers is not discernible from photographs, an upward-closing fold is properly described as an antiform, the opposite being a synform.



(a)

Fig. 4.33 Large shear zones must be painstakingly mapped out in the field, but are often well displayed on satellite images such as those of Landsat MSS band 5 shown here. In (a) the curvature and narrowing of pairs of ridges controlled by Archaean ironstones defines a large sinistral shear belt in



South India. In the Red Sea Hills of Sudan (b) an indistinct, NW-SE shear belt has sinistrally displaced an earlier WSW-ENE zone of shearing by about 30 km. Scale: 1 : 1 million. Source: author.

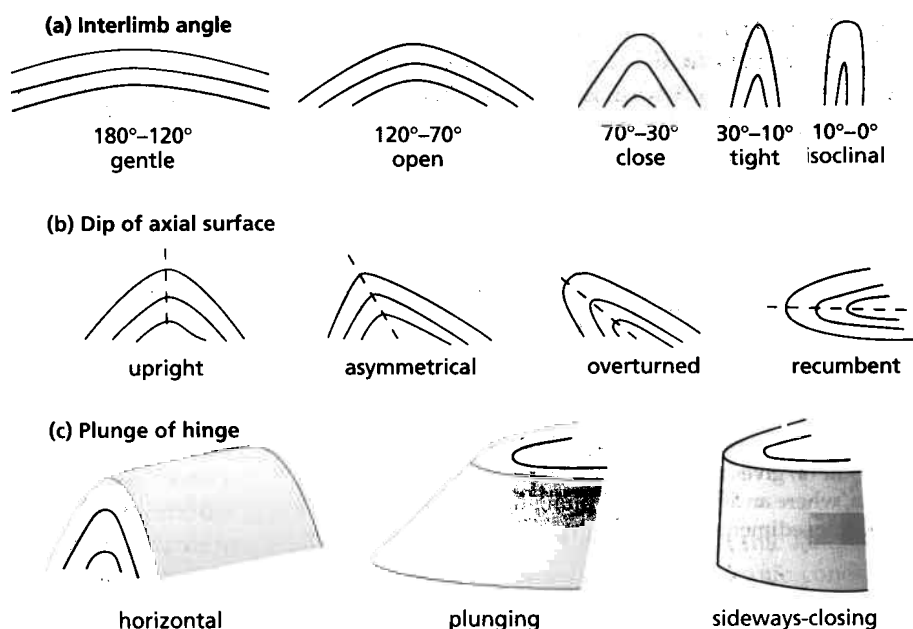


Fig. 4.34 The nomenclature of folds is based partly on three main features—the angle between the limbs of a fold, the plunge of its axis and the dip of its axial surface. Rarely more than these features can be seen on remotely sensed images.

Where a fold has a shallow plunge, closures are not common. The limbs crop out as dipping series of layers, taking forms appropriate to the angle of dip and the topography (Section 4.3.1). Detecting folding and locating fold

axes in such cases depends upon systematic observation of dip direction, using valley Vs (Fig. 4.35a,b). Like faults, however, folds do not propagate to infinity. At some point they die out, the strain being taken up by another

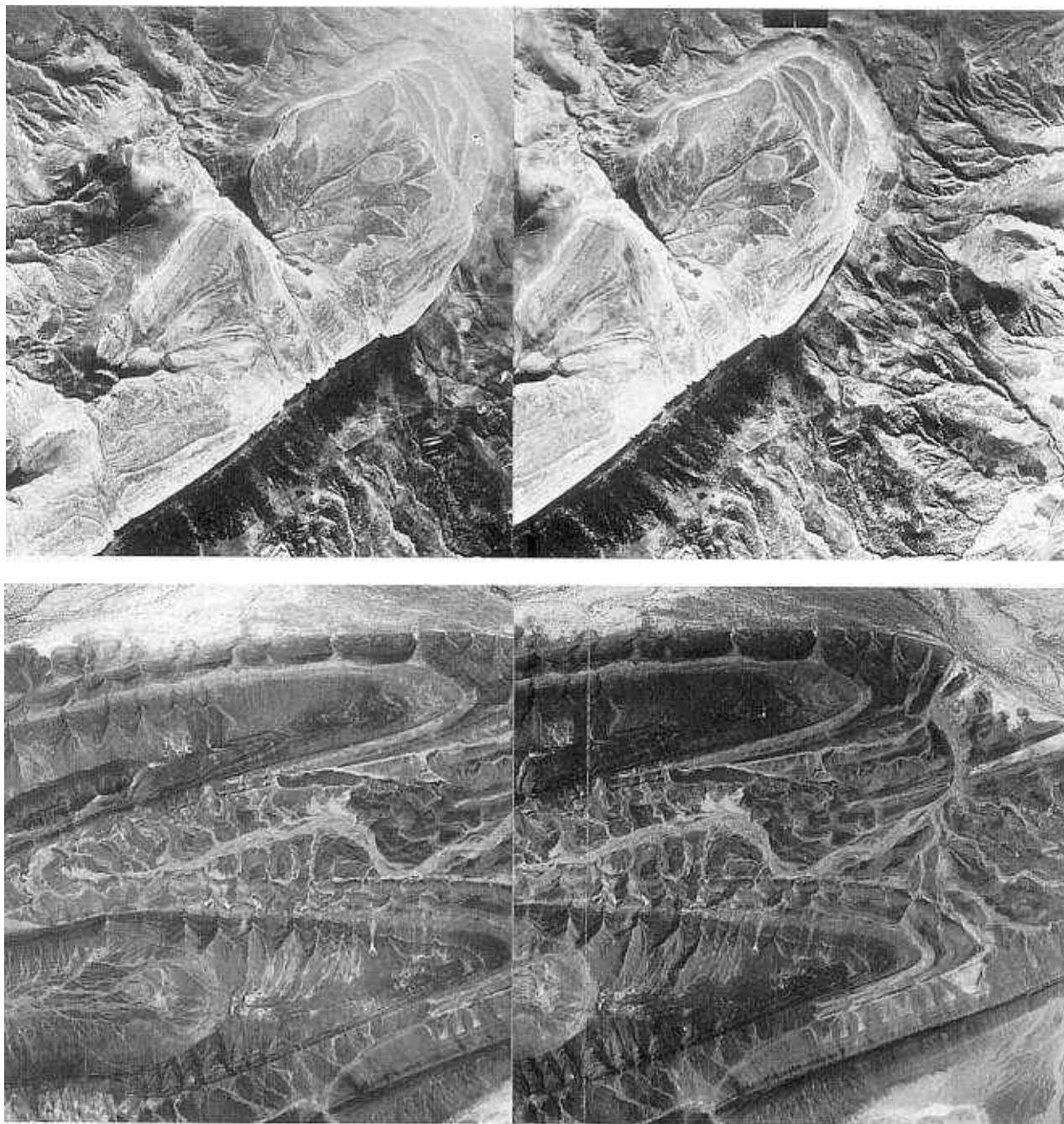


Fig. 4.35 Folds are the most spectacular features on remotely sensed images, particularly when they are well exposed. The stereopair (a) gives an example of inverted topography in a synform, where an asymmetrical, plunging synformal folding of Paleozoic sediments controls high ground. The nature of

the fold could be worked out from the varying directions of the axes of valley Vs by using one image. This would be possible on stereopair (b) too, which shows two plunging, metrical antiforms separated by a synform in Morocco.

Even the most regular folds have a plunge on their axes, and this may show as a closure of outcrops of some distinctive unit. The more steeply the axes plunge, the more obvious and more common the closures are (Fig. 4.35a,b). The forms taken by folds in the photograph are strongly controlled by the asymmetry of the dips on the limbs. The limbs of an upright

fold appear almost like mirror images of each other (Fig. 4.35b). In an extremely asymmetric fold the steeper limb is less controlled by topography than the shallow limb and adopts a more nearly linear form (Fig. 4.18). Turned folds present problems. The tighter and the more nearly recumbent they are, the more complex are the outcrops of their limbs (Fig. 4.35c). Where stereoscopic

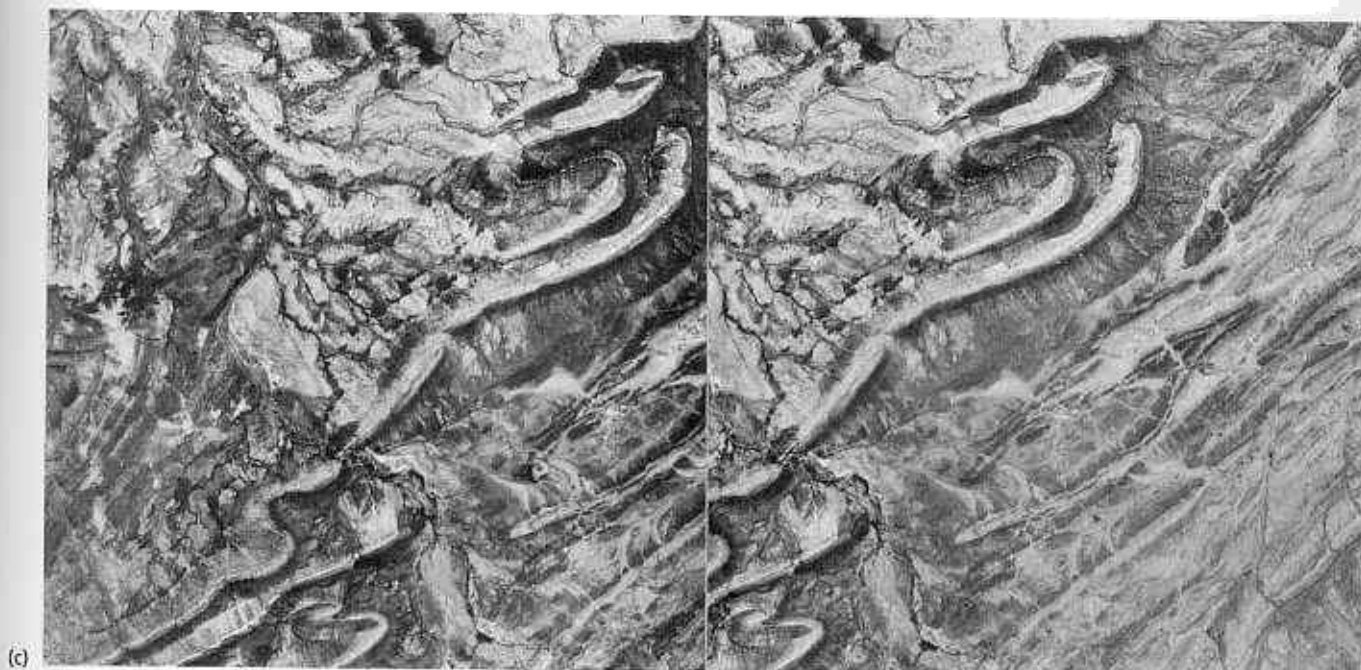


Fig. 4.35 Stereopair (c) contains several overturned folds in the Palaeozoic rocks of the Marathon Basin of Texas. The ridges are controlled by resistant carbonates and the sense of overturning is towards the north-west. The area is also

affected by thrusts, which disrupt some of the folds, and a few normal faults. Scales: (a and b) 1 : 40 000 and (c) 1 : 63 360.

Sources: (a) IGN ALG-1952-231-233 123 + 124, (b) IGN Maroc-009 424 + 425 and (c) NCIC Tex.1 A + B.

viewing is possible, the correct description of folds is relatively easy. For satellite images the process can be fraught with problems, the most common of which is confusion between antiforms and synforms (Plate 4.4). The most painstaking search for evidence of valley Vs is the only solution.

Among the most weird patterns that can be seen on remotely sensed images are those produced by the repeated folding of layered rocks. When a fold with one orientation is refolded by another with a different attitude a three-dimensional interference structure is produced. The distinctive patterns are produced when such structures are planed by erosion. They can adopt a bewildering variety of forms, but three types form the basic and most commonly seen end-members. Where upright folds are refolded by another set with a different axial trend, domes and basins result (Fig. 4.16c). Recumbent folds refolded by upright folds produce more or less mushroom-shaped interference patterns in horizontal section (Plate 4.5). The effect of strike-slip shear zones on pre-existing folds is often to transform them into hook-shaped structures (Fig. 4.33). The presence of any of these fold forms, and even odder ones too, is sufficient to raise the suspicion of polyphase deformation. Caution is needed however, as quite complex patterns can result purely by erosional means in the most innocuous terrains (Fig. 4.51).

4.5 Superficial deposits and constructional landforms

More even than destructional landforms, those resulting from the constructive action of volcanism and from sediment transport and deposition are enormously varied. So many factors are involved that the number of permutations and combinations is near infinite. Space does not permit a comprehensive treatment, merely an account of some common and easily recognized features. Most texts on volcanic, desert, glacial, fluvial and coastal geomorphology provide sufficient diversity of descriptions and photography for this brief account to be easily supplemented.

4.5.1 Volcanic landforms

The popular concept of an active volcanic landscape is that it is typified by symmetrical volcanic cones, each possessing a summit crater. This is only sometimes the case for areas dominated by viscous magmas, of intermediate to acid composition. These magmas cannot flow far from the vent from which they issue, and being unable to release their gas content easily, are prone to explosive activity. The result is that lavas and ejecta from a vent build up close to it as a layered conical structure

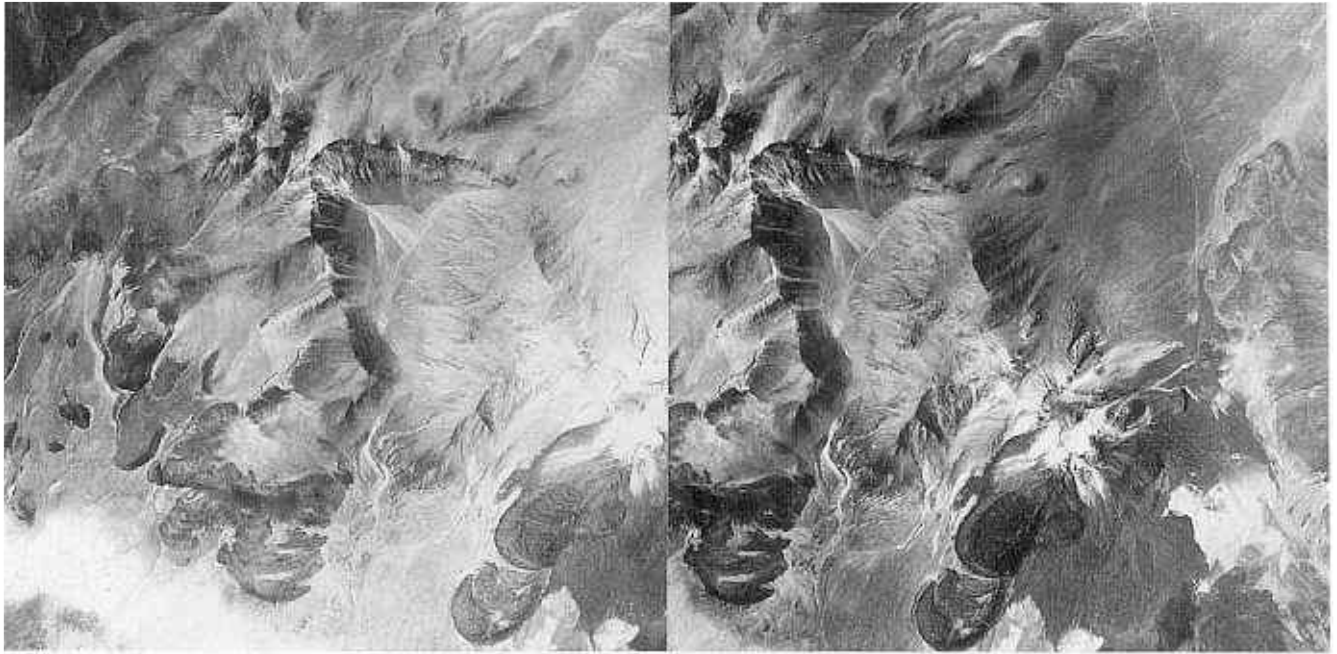


Fig. 4.36 As a result of a sideways directed explosion, this stratovolcano in the Andes of Chile has been breached to reveal its internal structure. The site of the blast is now occupied by a

massive dacitic resurgent dome. The flanks of the original volcano are irregular as a result of marginal flows of viscous lava. Scale: 1 : 60 000. Source: Instituto Geographico Militar, Chile.

termed a stratovolcano, which may be as much as several thousand metres high. The sides of a stratovolcano typically slope at angles around 30° . The vent itself is kept relatively clear of debris by repeated explosions, so

forming a crater. The strengthening of the volcanic edifice by lava flows ensures that the crater walls are steep and cliff-like. Figure 4.36 shows an example of the more typical, complex stratovolcano.

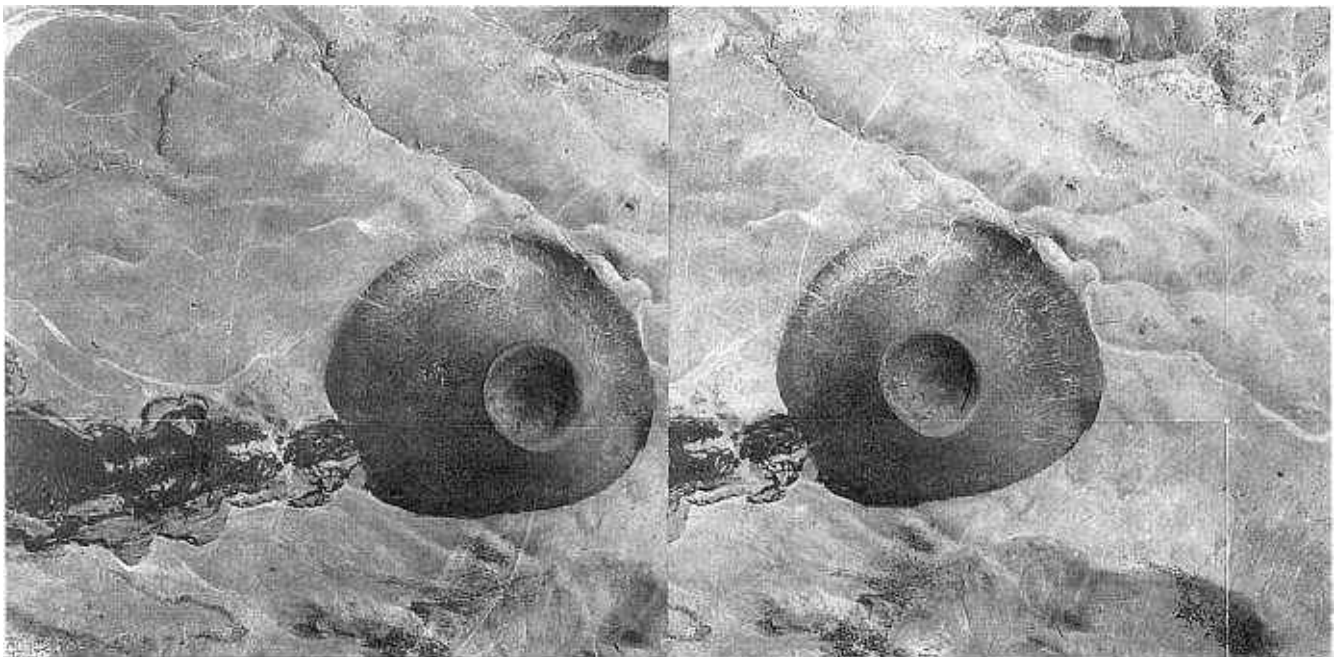


Fig. 4.37 Cinder cones, such as this typical example from Arizona, are closer to the popular concept of a volcano than most stratovolcanoes. The cone is symmetrical, despite the apparent distortion as a result of radial-relief displacement. The funnel shape of the crater, which results from the

instability of unconsolidated ash, is clear. In this case, explosive activity succeeded earlier eruption of moderately fluid lava, the typical morphology of which is obviously overstepped by the cone. Scale: 1 : 20 000. Source: ASCS BT 1952 + 1953.

Minor explosive activity ejects solid debris in the 5 mm to 1 m size range and also produces cones. The fragments are frequently full of gas bubbles and look like cinders. As they are loose they come to rest at a natural angle of repose depending on their size and shape. Such cinder cones therefore are characterized by having a smooth surface—at most scales involved in remote sensing—and uniform slopes. The slopes in craters are approximately the same as those on the cone sides, giving them a funnel shape. This means that they are easily distinguished from stratovolcanoes (Fig. 4.37). Cinder cones are rarely more than 300 m high and 1 km in diameter.

Basaltic magmas are much more fluid than those of intermediate to acid composition. They tend to release their gas content with a minimum of explosive activity, and travel long distances down shallow slopes. They rarely build steep cones, although there may be a minor basaltic component in stratovolcanoes. Two major kinds of basaltic landform are produced. The magmas may issue from fissures induced by crustal extension. Such fissure eruptions (Fig. 4.38a) typify basaltic volcanism at constructive plate margins. Where basaltic magmatism is controlled centrally, as in a mid-plate setting, lavas flow in all directions to build up vast domes with very

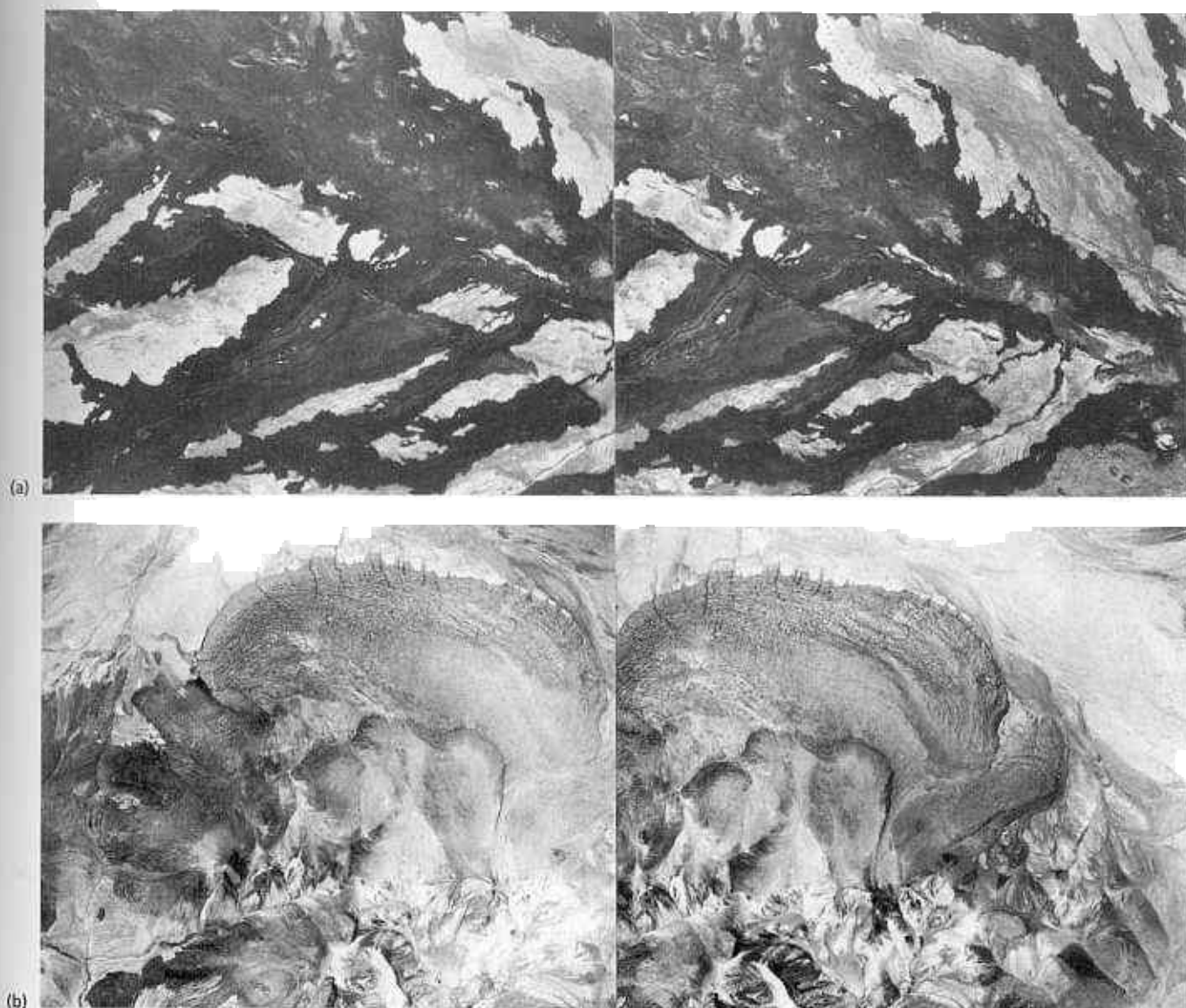


Fig. 4.38 On the flanks of Mauna Loa shield volcano on Hawaii, eruptions of fluid, gas-charged basaltic lavas emanated from long fissures (a). Small spatter cones are aligned along some of the fissures, as are cinder cones. The black dots on a pale background are collapse pits above a lava tube. Silicic lavas tend to be far more viscous than those of

basaltic composition. Stereopair (b) shows an excellent example from the Andes of Chile. It has not flowed far from the dissected stratovolcano that was its source. Its distinct lobate form shows ridges perpendicular to flow direction. Scales: (a) 1 : 41 000 and (b) 1 : 60 000. Sources: (a) NCIC Hawaii 7 and (b) Instituto Geografico Militar, Chile.

low slope angles—typically $5\text{--}10^\circ$. These shield volcanoes can be hundreds of kilometres across, as in the case of those in Hawaii.

Basaltic volcanism sometimes produces fire fountains as lava is forced violently to the surface by the pressure of rapidly rising magma. The lava forms droplets and globules, which are still partly molten when they land. This ensures that they stick together to form spatter cones (Fig. 4.38a). These tend to be associated with fissure eruptions, but may also form on a minor scale when lava erupts from the flanks of shield volcanoes or squirts through the crust of a major flow as a result of the pressure head involved. Spatter cones have very steep sides owing to rapid solidification. They are usually less than 20 m high.

Truly explosive activity with cinder and ash production is associated with basaltic activity only where water can enter the magma chamber. Basaltic cinder and ash cones are found in association with coastal volcanoes, and where volcanoes form beneath ice sheets, as in Iceland. In the last case truly bizarre geomorphological events, called *jökulhlaups* may occur, which involve glacial collapse, catastrophic floods and mudflows as well as the formation of volcanic edifices.

Undissected lava flows have very distinctive appearances resulting from the process of flow itself. The most obvious is the formation of lobate margins, where the cool end of a flow is highly viscous and spreads slowly into topographic depressions. The solid crust buckles and breaks into blocks, which can produce *levée*-like ridges at the sides and ends of flows (Fig. 4.37). In viscous lavas, the formation of ridges perpendicular to flow direction extends right across flows (Fig. 4.38b). Such lavas flow only down steep slopes, and usually extend only short distances. Basaltic flows travel much further, and tend to develop flow lines parallel to flow direction (Fig. 4.38a). Sometimes late-stage lava movement is beneath an already solidified crust. The resulting lava tubes can collapse to form distinctive sinuous depressions on the flow surface (Fig. 4.38a).

In areas of repeated lava eruptions distinguishing between flows of different ages is sometimes a relatively simple matter. The stratigraphic rules of superposition and overlap generally can be applied. Moreover, the older a flow the greater the tendency for its surface roughness to be degraded by soil formation, and for it to support vegetation (Fig. 4.39).

The issue of fine-grained pyroclastic material from volcanoes prone to explosive activity produces a blanketing effect over older landforms. The resulting tuffs smooth out the topography (Fig. 4.40) and kill the local vegetation. They are, however, prone to rapid erosion, and are soon dissected. The most spectacular results of explosive activity relate to the flow of high-density, incandescent mixtures of ash and gases. These *nuées*

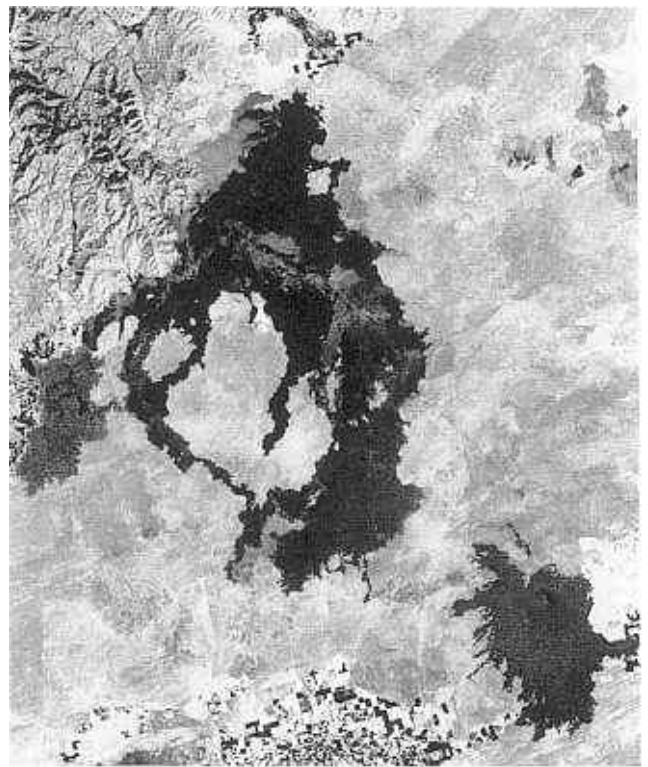


Fig. 4.39 Lava flows of different ages are distinguished on this Landsat MSS band 7 image of part of Idaho by their different reflectances. The youngest flows are black and successively older flows are increasingly reflective as a result of being masked by blown sediment and colonization by vegetation. The flows emerged from a NW–SE fissure, which can be seen faintly near the centre of the largest area of lavas. Scale: 1 : 1 million. Source: courtesy of Dave Rothery, Open University.

ardentes are emitted from volcanoes with such viscous and gas-charged magmas that they cannot easily be extruded. Gas pressure builds up in the magma chamber until the entire contents are belched out in a single event. The dense mixture of hot ash and gas has a low viscosity, so that its immense gravitational energy causes the mixture to travel at velocities in excess of 100 km per hour. It therefore can flow over and around large objects in its path to swamp vast areas and will eventually solidify as an ignimbrite (Section 4.2.2). The high energy of *nuées ardentes* results in deep fluting of the ignimbrite surface, parallel to the direction of flow (Fig. 4.40). Recent ignimbrites are thus easy to identify, even from satellite altitudes.

Stratovolcanoes, being formed by explosive magmatism, are prone to catastrophe. The expulsion of the contents of a magma chamber during a *nuée ardente* event leaves a cavity at depth, into which the superstructure may collapse. This is accomplished by annular systems of faults defining a caldera (Fig. 4.41a). The resulting circular depression is occasionally filled with

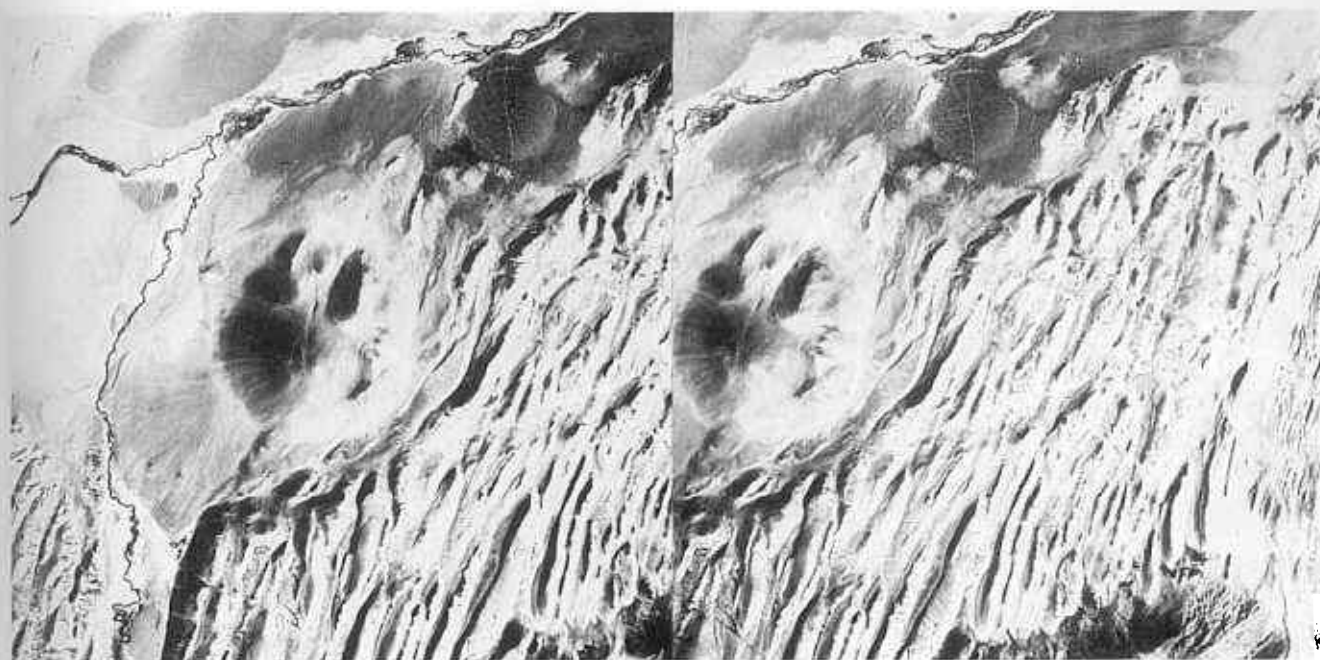
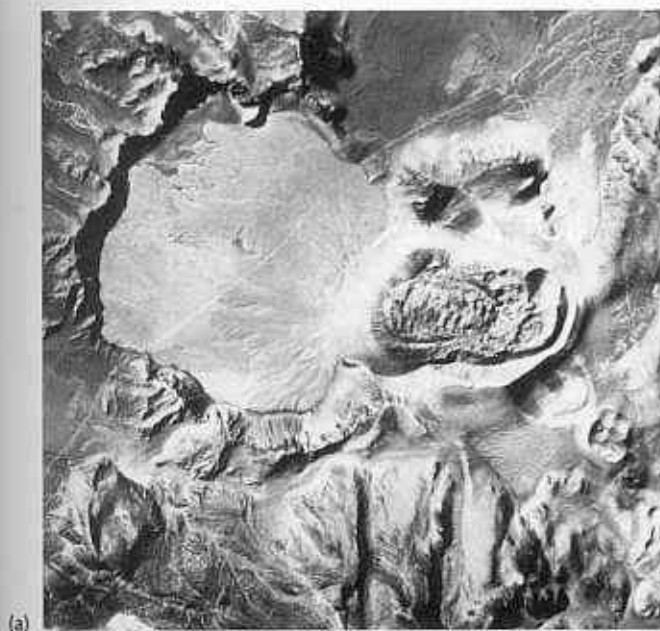


Fig. 4.40 The most characteristic feature of the huge ignimbrite flows of the central Andes is their deep fluting. Associated with this flow in Argentina is a large dacitic dome. Scale: 1 : 60 000. Source: Instituto Geografico Militar, Argentina.



(a)

Fig. 4.41 The caldera in the Argentinean Andes shown in (a) is partly filled with later silicic lava flows from a centre now occupied by a lava dome. The dome developed on the fault bounding the caldera. The Landsat MSS band 7 image in (b) shows the volcano Socompa in Chile, which underwent a catastrophic lateral blast that dwarfs the famous Mount St Helens event of May 1980. Most of the surface in the lower part of the image consists of a debris avalanche, formerly



the southern flank of the volcano. The complex tonal and textural patterns in the debris can be used to identify the original location of the material in the volcanic superstructure and to reconstruct the mechanics of the avalanche. Scales: (a) 1 : 60 000 and (b) 1 : 400 000. Sources: (a) Instituto Geografico Militar, Argentina and (b) courtesy of Dave Rothery, Open University.

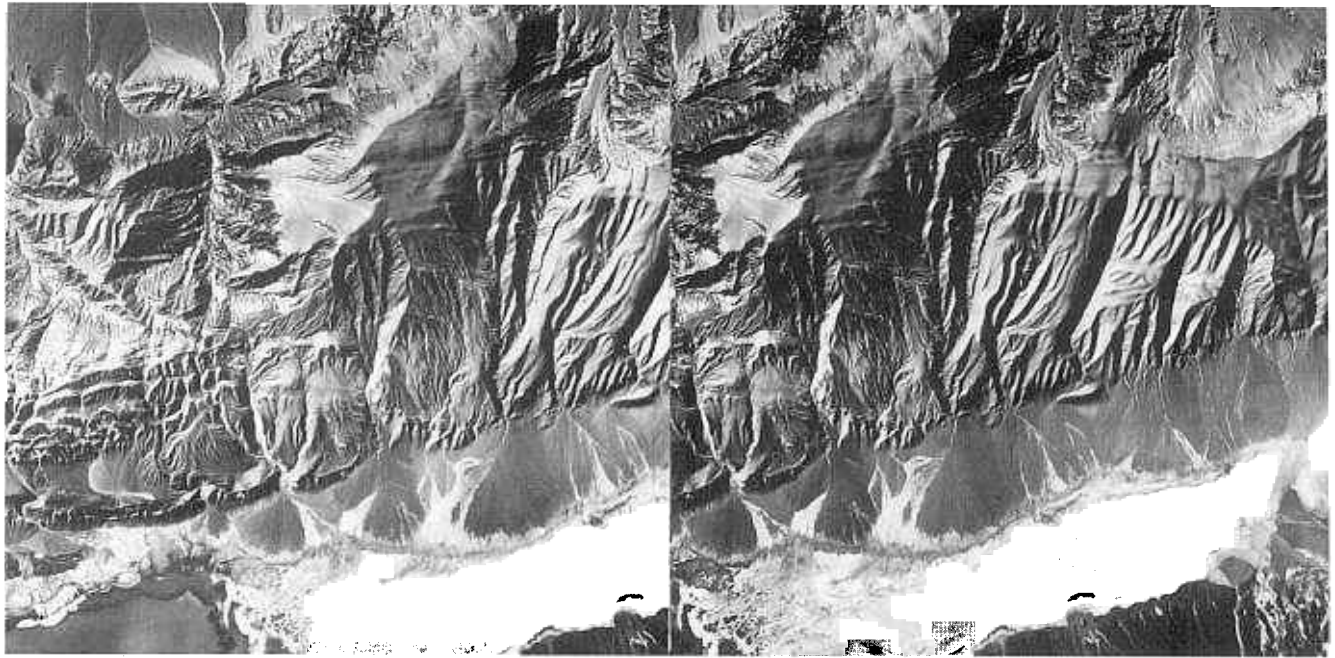


Fig. 4.42 Alluvial fans typically develop at the exits of intermittent streams draining arid mountainous regions. In this part of the Andes of western Argentina, steeply dipping Palaeozoic strata have received a thin veneer of volcanic debris from a massive ignimbrite event resulting

in a peculiar smoothing of the older topography. Erosion of the veneer has not long been underway, and is marked by localized, complex systems of sharp rills. Scale: 1 : 60 000. Source: Instituto Geografico Militar, Argentina.

a lake. Continued magmatism may exploit the caldera to form minor cones along the annular faults, and central domes of resurgent, viscous magma (Fig. 4.40). Instead of resulting in nuée ardente activity, build up of pressure within a volcano may cause the flanks to bulge alarmingly. If the bulge develops unstable slopes, the volcano flank may collapse to form a massive debris flow (Fig. 4.41b). The release of pressure on the flank can then allow the accumulated gas-rich magma to explode sideways, breaching the entire superstructure. This was the process associated with the May 1980 eruption of Mount St Helens in north-western USA.

Many stratovolcanoes reach high elevations, and even in equatorial latitudes may be capped with snow and ice. This is a hazardous association. A combination of ice-bound ash and heat associated with episodes of activity can produce sudden melting and the formation of mud avalanches. These can assume nearly the same monumental proportions as ignimbrite and volcanic collapse events. The mud flows in much the same fashion as lava, usually faster and further. The energy of these volcanic mudflows is high enough to sweep all before them, leaving sinuous scars through vegetation, and blocking local drainage when they come to rest. Their process of solidification involves a fall in the velocity of turbulent flow, below which the mud ceases to behave as a liquid and assumes solid proportions. It sets, and slowly the water content seeps away to leave a resistant mass of

ash and debris, which has many of the morphological features of a lava flow. Being soft, however, they are soon eroded away.

4.5.2 Fluvial and lacustrine landforms

Deposition by flowing water to form constructional landforms results when flow velocity falls below that needed to retain sediment in suspension or to move it along the stream bed by bouncing or rolling.

High-energy streams issuing from uplands carry abundant coarse particles. Once the slope changes as the streams emerge from constricted valleys, this coarse load is quickly dumped to form low cones with their apices at the mouths of the mountain valleys. The streams are able to flow in any radial direction on these cones, and so split into numerous divergent channels. This gives rise to alluvial fans, which typify the flanks of uplands in semiarid terrains (Fig. 4.42). Stream flow is sporadic and reaches a maximum during storms or rapid snowmelt at high altitude. The constricted flow spreads at the valley mouth to form sheet floods which deposit and redistribute debris uniformly across the fans. The highly porous nature of the fans ensures that a high proportion of flow infiltrates to produce local groundwater supplies. Only where vegetation is unable to bind the debris and stabilize stream courses do alluvial fans in humid terrains develop on steep slopes as screes.

The lower reaches of a mature or old-aged stream have low slope angles. When the stream leaves its course during floods most of the suspended load is deposited on the flood plain as alluvium. Streams flowing across almost flat ground have an irresistible tendency to form meanders. The protracted migration of meanders across flood plains reworks the alluvium deposited there to produce several distinctive structures. The inside of each bend is the site for deposition, the outside being eroded. Meanders migrate towards their outside banks as well as downstream. This produces a system of low, arcuate features, known as meander scrolls, which represent

deposition on the insides of bends (Fig. 4.43a). Sometimes meanders curve back upon themselves to such a degree that the narrow ground separating parts of a bend is cut through. The bend is cut off to form a stagnant oxbow lake (Fig. 4.43a). Features such as these typify flood plains where the alluvium is sufficiently well-bound by vegetation or because of its fineness to support discrete stream courses. In semiarid terrains or in flat valleys filled with poorly bound alluvium, stream courses are ephemeral. Flow may change from one course to another repeatedly, thereby forming a braided stream system (Fig. 4.43b).

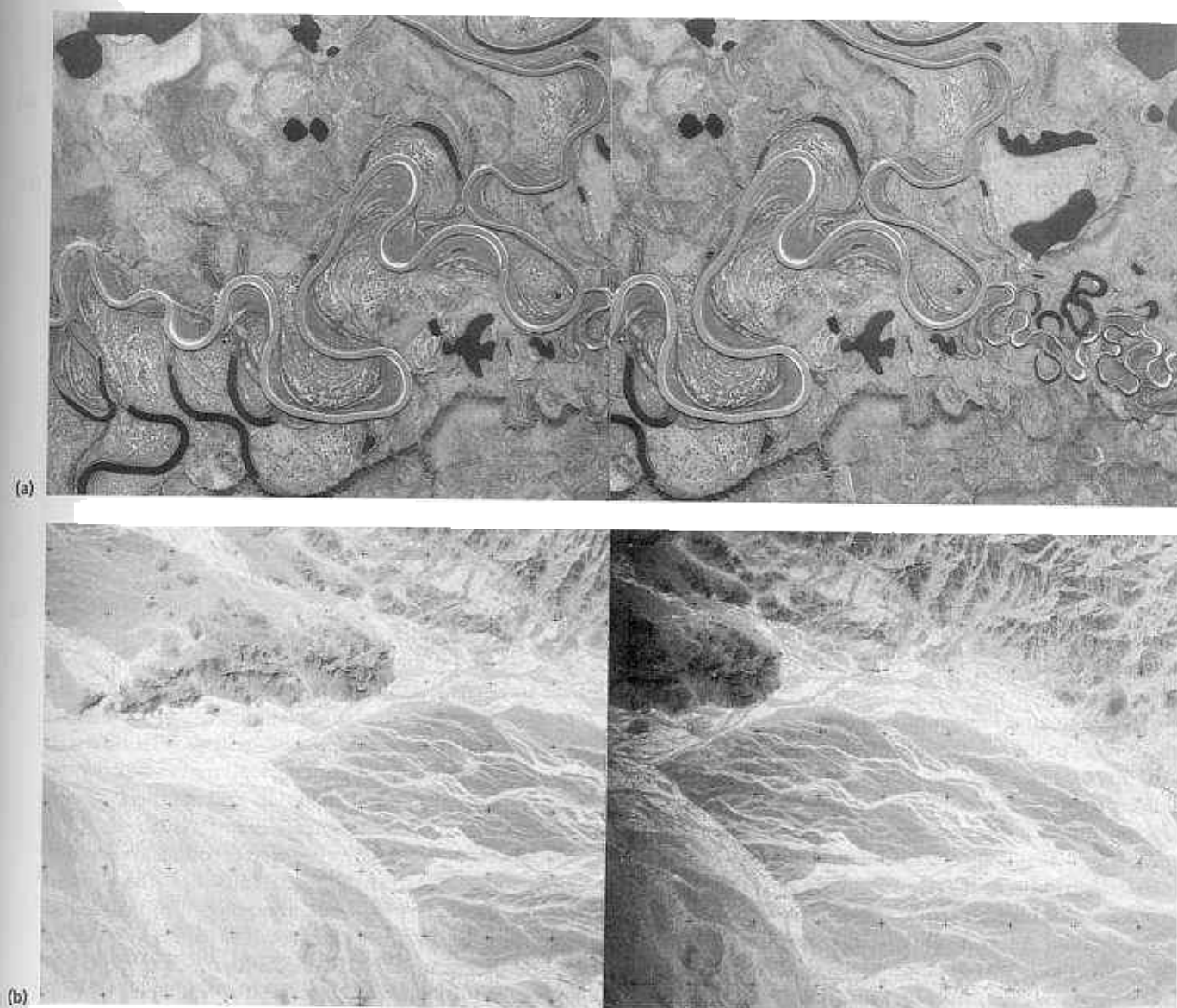


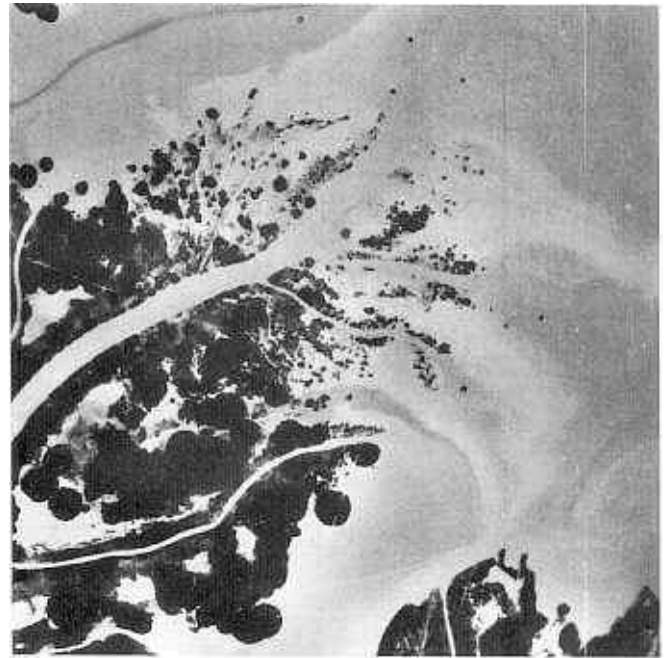
Fig. 4.43 Continuous flow of streams over nearly featureless plains of unconsolidated sediments results in migrating meanders. In the example shown from the Canadian Arctic (a) traces of former meanders are preserved as meander scrolls on the insides of bends, and abandoned channels in the form

of oxbow lakes. Braided stream systems form on similar substrates when flow is intermittent. As in (b), they are best developed in semiarid terrains. Scales: (a) 1 : 70 000 and (b) unknown. Sources: (a) EMRC A13139 82 + 83, copyright HM the Queen in Right of Canada (1951), and (b) unknown.



(a)

Fig. 4.44 Deltas formed by shifting distributary channels take on the classic arcuate form, typified by that shown in (a). It results from the entry of a large stream into a lake in New York. Where sediment is deposited in standing water by a number of fixed distributaries, most material is deposited near the



channels and builds up deltas of the birdsfoot kind (b). This system is part of the Mississippi delta in Louisiana. Scales: (a) 1 : 20 000 and (b) 1 : 60 000. Sources: (a) ASCS ARU-IEE 74 + 75 and (b) NCIC La.8.

Changes in erosional base-level in a flood plain result in the stream system cutting into earlier alluvial deposits. This produces terraces of flat ground above the active flood plain, on which old fluvial features may be preserved, and into which minor drainage may have been incised. It is often possible to recognize several terrace levels. Those with coincident heights on either side of the flood plain are the result of rejuvenation. Those that have different heights either side are the products of slow downward erosion and migration back and forth of the stream system.

Where streams flow into bodies of standing water, either lakes or the sea, the flow is halted and all sediment is deposited. The result is a delta. If the delta develops shifting distributary channels, similar to those in an alluvial fan, then it adopts an arcuate shape as a low-angled cone (Fig. 4.44a). Where persistent channels are maintained deposition is from these and the delta propagates as a branching birdfoot delta (Fig. 4.44b).

Deposition in lakes, other than that associated with deltas, is in the form of horizontal layers or in beaches. Drainage of a lake results in almost perfectly flat topography, interrupted only by the relics of former islands. Old lake beds can be distinguished from abandoned flood plains by the absence of meander scrolls and oxbow lakes, and the presence of curvilinear features marking the sites of old beaches (see Section 4.5.4).

4.5.3 Aeolian landforms

The action of wind is not restricted to deserts, but can mould constructional landforms wherever sediments of a size small enough to be transported in air are not bound together. They may contribute to both coastal and glacial landscapes. The physics of wind and its means of transporting particles are complex and sensitive to many subtle factors. The dominant process is the efficient sorting of grain sizes into the bouncing and rolling load and that carried in suspension. For a particular wind speed the ranges of grain sizes present in both loads is very narrow. Aeolian deposits are therefore very well sorted. The rolling and bouncing load moves much more slowly than suspended material, and so the two size ranges are deposited in completely different areas.

Dunes result from the deposition of rolling and bouncing particles. Their form is determined by the amount of sand available in an area, the amount of vegetation cover and by the constancy of the wind. The classic migrating, horn-shaped dune or barchan (Fig. 4.45b) forms where sand supply is low, little vegetation is available to halt the movement of sand and wind direction is roughly constant. The horns point downwind. The lee slope, where deposition takes place, is steeper than the windward slope up which grains are blown. Barchans migrate slowly down wind. Where vegetation is able

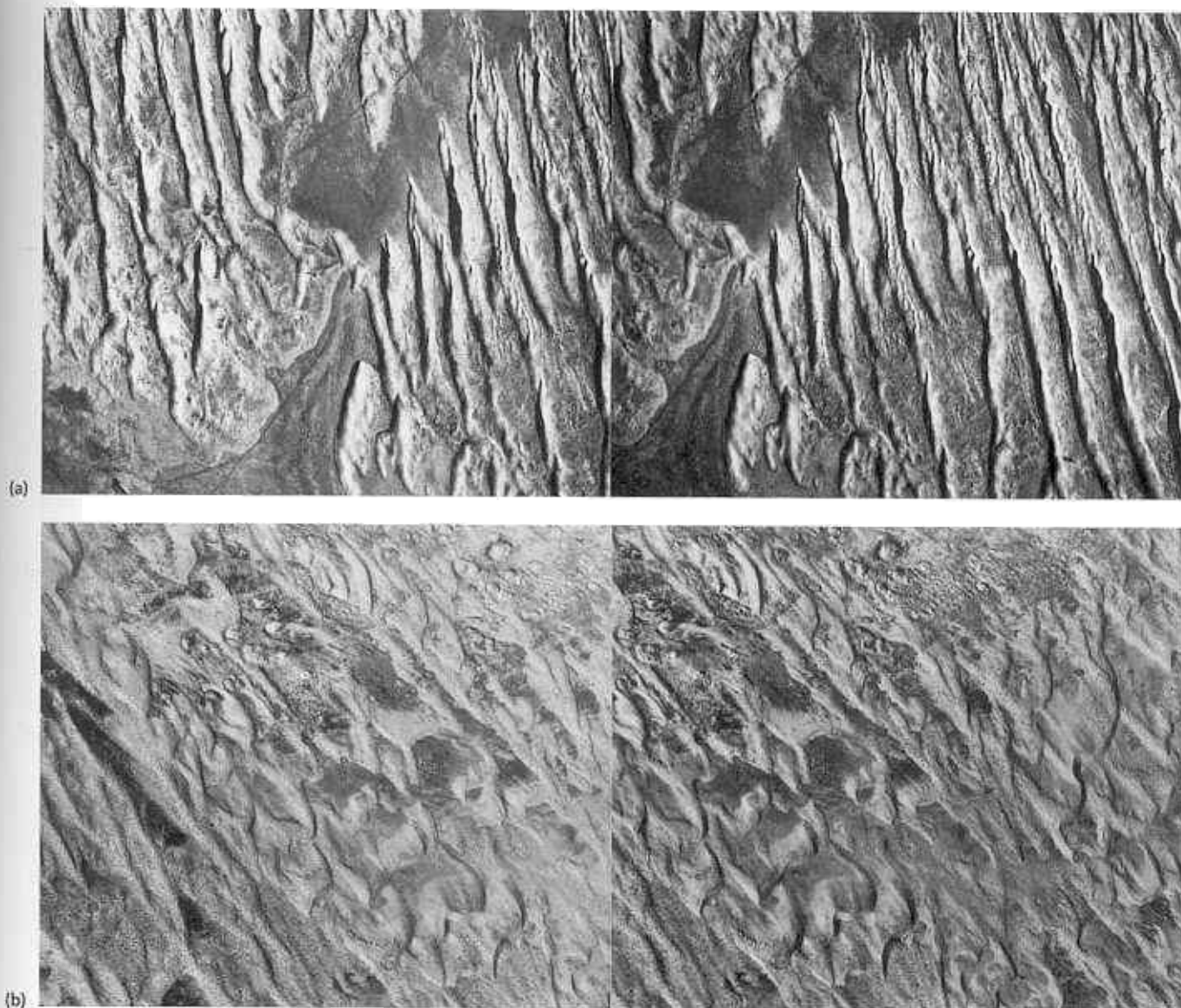


Fig. 4.45 An abundant supply of sand, dry conditions and strong prevailing winds favour the linear or seif dunes shown from an area in the Middle East by (a). The major dunes are parallel to the prevailing wind and on their crests have small migratory dunes, the horns of which point downwind.

Where sand supply is restricted, as on the Kaibab Plateau of Arizona (b), parabolic, transverse and barchan dunes dominate. Scales: (a) 1 : 20 000 and (b) 1 : 40 000.

Sources: (a) CSL Series B 9 + 10 and (b) NCIC Ariz.7 B + C.

to bind sand as it accumulates, dunes with the same overall shape as barchans develop, but instead of the horns pointing downwind they are fixed by vegetation and point windwards.

An abundance of sand and a lack of vegetation encourage the formation of large dunes with crests that line up perpendicular to the prevailing wind. These transverse dunes are typical of the world's largest sandy deserts. Where wind velocities are high, however, sand becomes stabilized in longitudinal or seif dunes running parallel to the wind (Fig. 4.45a). These can extend for hundreds of kilometres, and are so large and stable that they can be

preserved after climate has become more humid, when the dunes become vegetated.

In deserts subject to seasonal fluctuations in wind direction, the dunes take on shapes conforming to the interfering effects of several directions of sand transport. They may have a random pattern, but sometimes develop distinctive star shapes.

The material carried in suspension by wind travels long distances until it reaches climatic regions with low wind speeds. These usually are associated with the high pressure systems over continental interiors. Settling of windborne dust produces a peculiar sediment known as

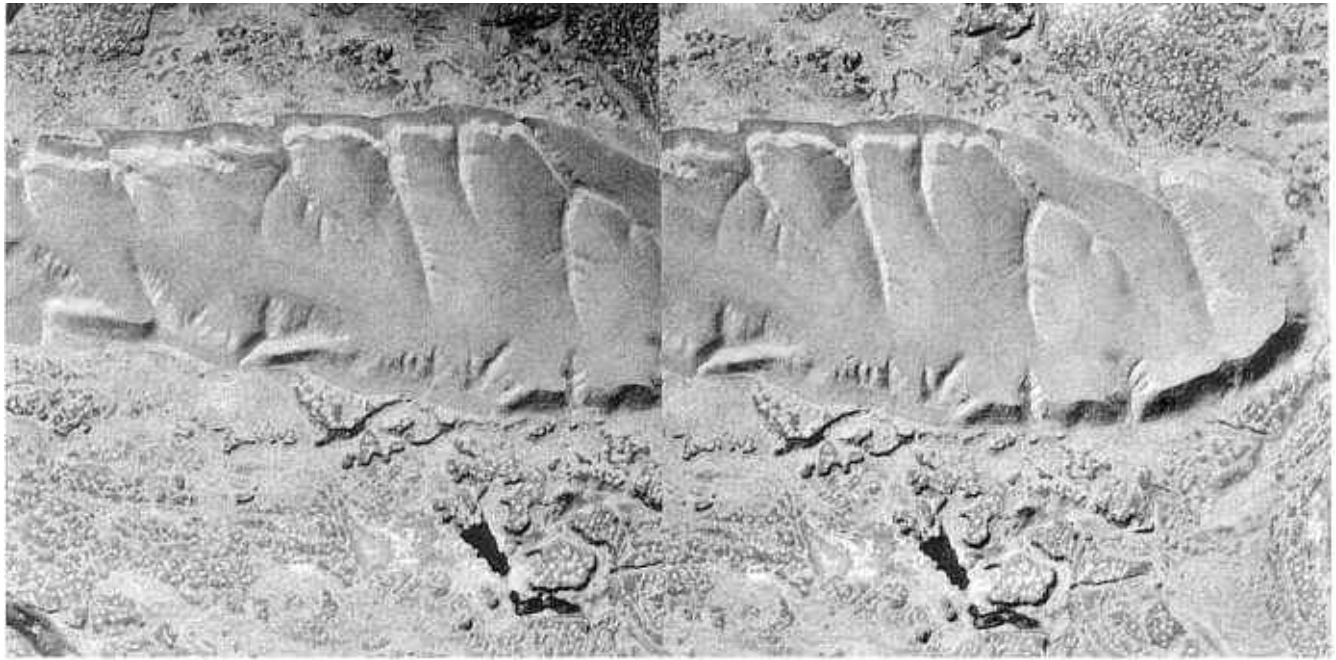


Fig. 4.46 The very finest material stripped from dry, bare sediments by the wind is carried long distances in suspension to be deposited as homogeneous mantles of loess. Stream erosion of loess produces a characteristic pinnate, or feathered drainage pattern well displayed by this stereopair of an area of

Washington. The peculiar texture in the area stripped of loess, known as scabland, is a product of scouring and giant ripple formation by powerful stream action in glacial-lake overflow channels. Scale: 1 : 20 000. Source: ASCS AAP-1G 123 + 124.

loess. It is exceptionally well sorted and homogeneous, and frequently lacks bedding. One consequence of this homogeneity and the process of deposition is that loess has a tendency to break along irregular vertical surfaces. This imparts a distinctive form of drainage pattern to loess-mantled areas in North America and Eurasia. They are dominated by dendritic networks of vertical-sided, U-shaped valleys and gullies. Each stream network looks remarkably like a feather, so they are described as pinnate (Fig. 4.46). Most of the largest areas of loess stem from ablation of fine dust from glacial deposits during the Pleistocene, when wind systems controlled by high pressure over the ice masses blew from west to east in the Northern Hemisphere.

4.5.4 Glacial landforms

Constructional landforms resulting from glacial periods present the greatest variety of all. Not only ice and its load of gouged-up sediment were involved, but also the meltwater and wind, which reworked the sediment, and the freezing temperatures, which imposed various structures upon and within it. It is impossible to give a comprehensive account of all the possibilities here, only those distinctive enough to identify terrains dominated by glacial and related deposits. A thorough analysis of photographs of such areas requires reference to specialized texts.

The material deposited solely by melting ice during its retreat is given the catch-all name till, which covers a multitude of sins. In an area of upland glaciation, till takes distinctive forms related to various kinds of moraine. Ice sheets in lowland areas deposit till in two distinct forms. At the ice front where supply is balanced by melting and ablation, debris is supplied continually. It is simply dumped and accumulates in a zone of exceedingly dirty ice. This forms an end moraine of one of several varieties. They may mark the ultimate extent of the ice sheet, or temporary halts during its retreat. Those associated with continental glaciation may extend as more or less continuous and sinuous lines for thousands of kilometres. They are characterized by a narrow belt of hummocky topography in an otherwise almost featureless plain. As till has a high content of fines, moraines are poorly drained and often have small lakes preserved in irregular depressions (Figs 4.47a and 4.48). These result either from the irregular nature of the moraine or from the melting of buried ice masses and consequent collapse. Collapse structures of this kind are known as kettle holes.

Till deposited at the base of a moving glacier has often been planed almost flat. Apart from sporadic kettle holes it presents a depressingly boring prospect. Its uniformity, impermeability and low relief ensures poor drainage and the establishment of complex dendritic and deranged patterns. Subtle variations in relief and

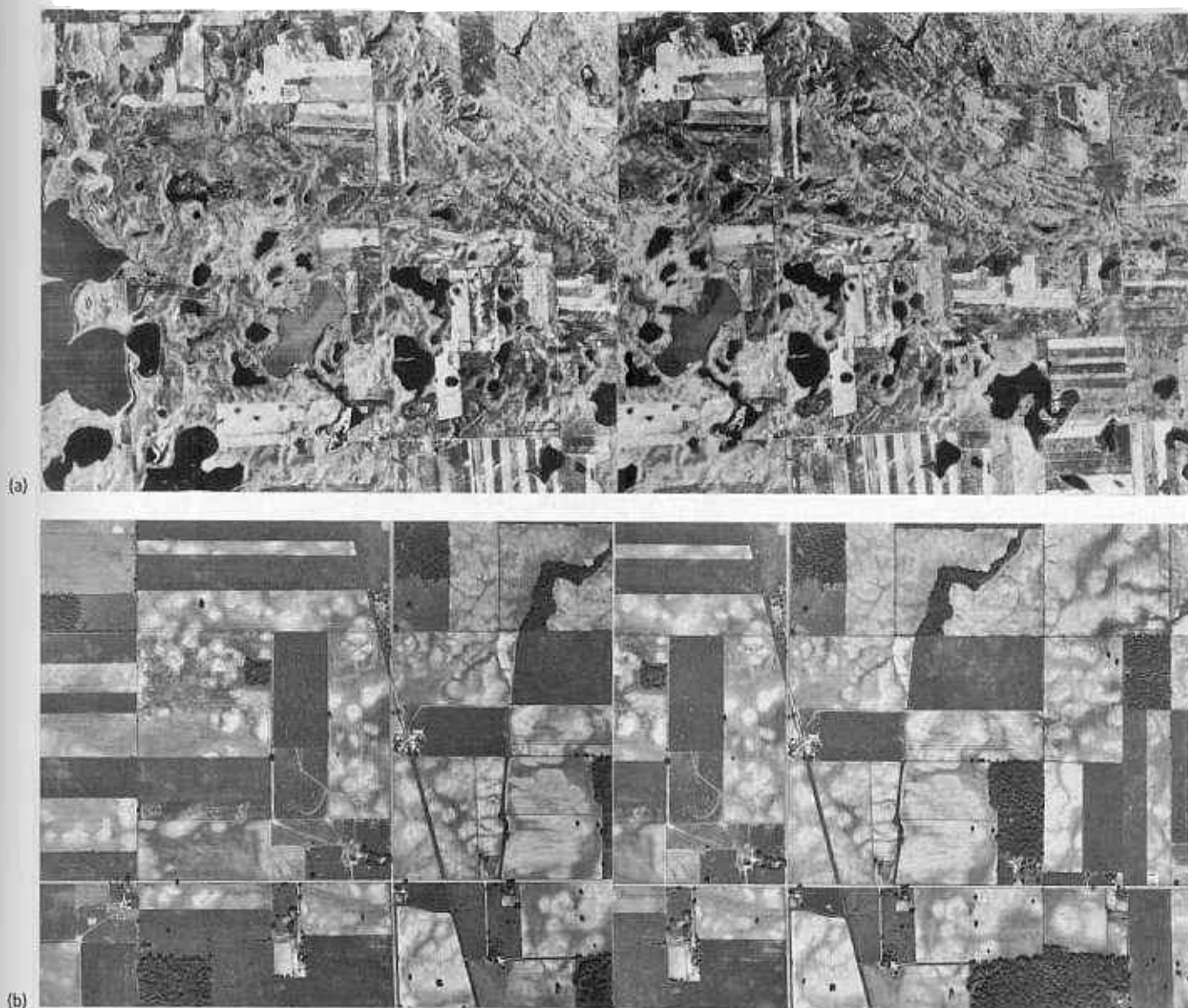


Fig. 4.47 Glacial till deposited at the front of a continental glacier was riddled with pockets of 'dead' ice. Their melting produces an irregular, hummocky topography on ill-drained, boulder clay. In the case of such end moraines in North Dakota (a), the depressions become filled with small lakes, and drainage is deranged. Also shown are areas of till with parallel linear ridges, which represent deposition beneath a moving ice sheet. Well within the area mantled by ice, basal till is

planed to an almost flat surface. Drainage takes no particular direction and rainfall eventually seeps into the till. The area of basal till in Indiana, shown in (b), is characterized by a mottled appearance owing to variations in soil moisture controlled by the irregular but subdued topography. Scales: (a) 1 : 60 000 and (b) 1 : 20 000. Sources: (a) NCIC N.Dak.2 B + C and (b) ASCS BWI-3BB 95 + 96.

permeability help produce a mottled texture of vegetation, particularly in spring (Fig. 4.47b). The grinding monotony of basal till is relieved by features produced by the once overriding ice mass. The most impressive of these are rounded, oval-shaped knolls known as drumlins (Fig. 4.48). The typical 'basket of eggs' topography of drumlin country results from irregularities in ice flow, brought on perhaps by obstructions beneath or by confluence of several sheets. The long axes of drumlins are ambiguous indicators of ice-flow direction.

Melting of the Pleistocene continental glaciers provided huge volumes of water. Where this meltwater flowed within the ice, meandering tunnels formed. In them the glacial debris was sorted and transported. Some of it was deposited in the base of the tunnels, to result in sinuous ridges of well-drained sand and gravel, termed eskers, when the ice melted (Fig. 4.49). Somewhat similar tracts of well-drained sandy terrain formed from deposits in ice-margin streams, where the glacier abutted areas of high relief. These well drained kame

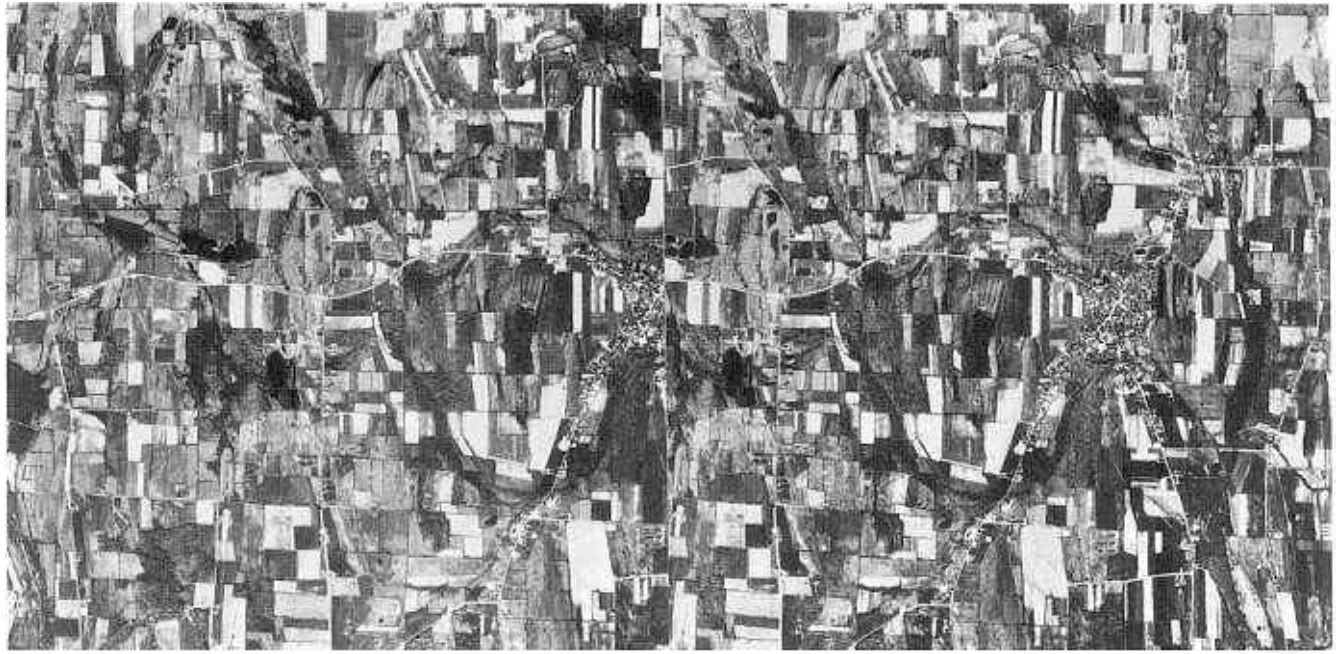


Fig. 4.48 The masking effect of regular fields in this stereopair of part of New York completely obscures the topography until the scene is viewed stereoptically. Then a series of drumlins becomes quite obvious, which has the typical 'basket-of-eggs' landscape. Scale: 1 : 60 000. Source: NCIC N.Y.3 A + B.

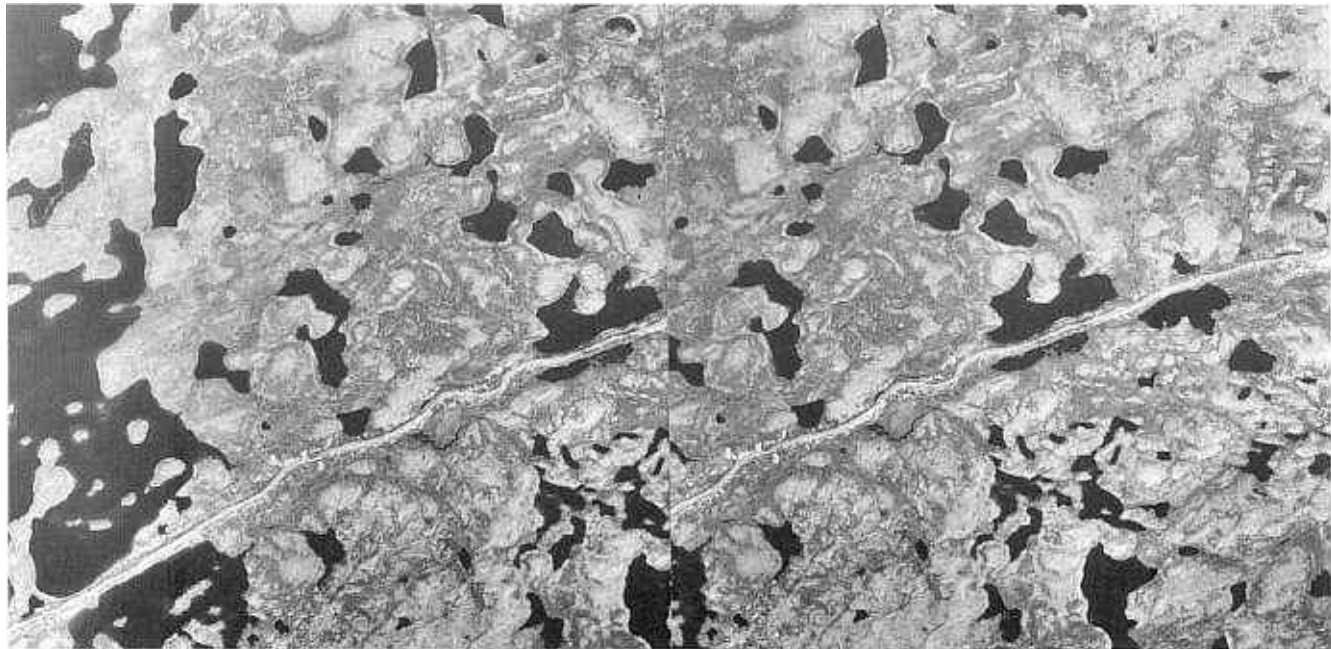


Fig. 4.49 The dominant material beneath this area of northern Canada is till deposited in an area of 'dead' ice, with its typical knob and kettle topography. Running across this ill-drained landscape is a sharp ridge of light-coloured, well-drained

sand, lacking vegetation. It is an esker deposited in a subglacial melt stream. Scale: 1 : 54 000. Source: EMRC A14887 102 + 103, copyright HM the Queen in Right of Canada (1955).

terraces have a contrasted vegetation to the impervious tills that they cross.

In front of the ice margin water flowing from the top of the ice or from tunnels within it was heavily charged

with sediment. The ice sheet, often up to a kilometre in thickness, formed prominent relief. The high energy streams flowing from it slowed rapidly and deposited their sediment load to form outwash deposits. They are

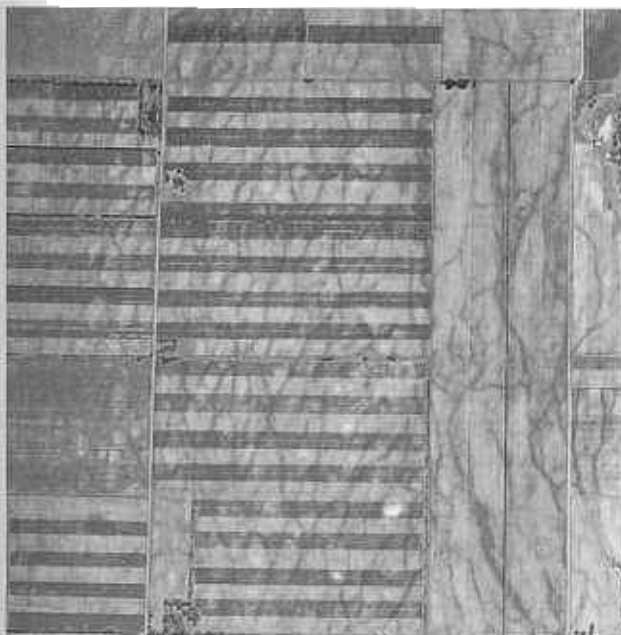


Fig. 4.50 Glacial outwash plains formed in front of continental glaciers, such as this one in Wisconsin, are nearly featureless on the ground. However, from the air the variations in soil moisture owing to different porosities shows up the relict patterns of the former braided distributary channels. Scale: 1 : 27 000. Source: ASCS BRT-9 V 38.

broad belts of sandy sediment that become progressively finer away from the ice front. In detail they are fans which coalesce, and on which braided streams flowed. This basic structure is still easily recognized in many cases (Fig. 4.50). Where outwash engulfed patches of ice left behind by glacial retreat, kettle holes formed, so that a pitted surface is sometimes seen. As they are so porous, outwash deposits are well drained and now have widely spaced surface drainage.

Pre-glacial drainage may have been dammed temporarily by ice so that large glacial lakes formed. The deposits resulting from such lakes are identical in most respects to those from any other drained lake. They are even more flat and tedious than ground moraine. Drained glacial lakes may even preserve relics of deltas from the melt-water streams that fed them. The sudden removal of the ice dam during glacial retreat released huge floods, which carved out deep valleys. Once the lake had drained, these valleys became occupied by streams too small to have been capable of forming them. The valleys have a typical J-shaped cross-section, revealing that they were in fact the channels of these short-lived torrents. They are sometimes characterized by huge potholes and giant ripples producing scabland (Fig. 4.46). Episodic drainage of glacial lakes resulted in successive shore lines being preserved as more or less parallel beach ridges (Fig. 4.51).

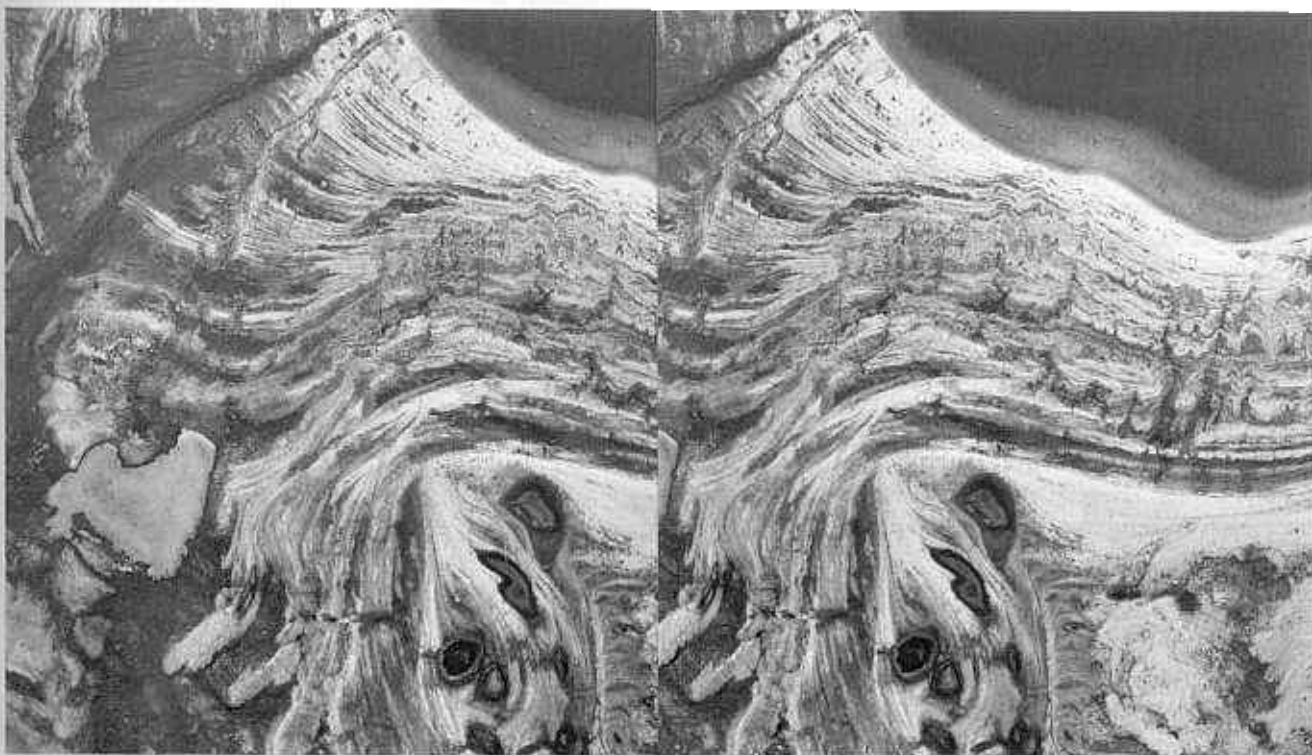


Fig. 4.51 At first sight the curvilinear features in this area of northern Canada are reminiscent of a complex folded terrain. Stereoscopic viewing reveals that the features are low ridges at successively higher elevations that were deposited as beaches,

pits and bars during a period when the glacially depressed crust was slowly rebounding isostatically. Scale: 1 : 36 000. Source: EMRC A16347 29 + 30, copyright HM the Queen in Right of Canada (1958).

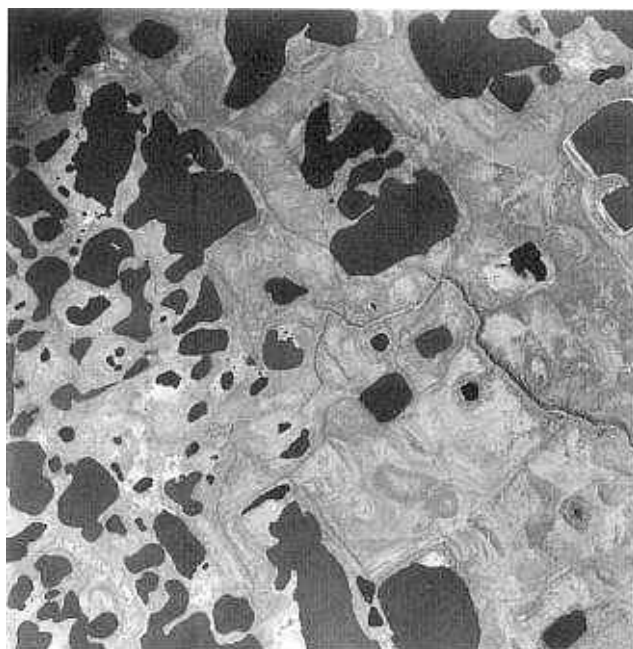


Fig. 4.52 Where unconsolidated sediments in northern Canada and other boreal areas are permanently frozen at depth, the seasonal freezing and thawing of the upper soil layer produces regular patterns. This is accomplished by the volume changes involved in ice formation and melting, which slowly circulates and sorts grains of different grain sizes. Another contributory factor is the formation of ice lenses and doming of the surface, followed by melting and collapse. Scale: 1 : 36 000. Source: EMRC A16347 30, copyright HM the Queen in Right of Canada (1951).

The unifying feature of all glaciated terrains is that they were subject to intense cold for long periods. In common with modern boreal areas they were prone to permafrost. Among the many effects of permanently frozen ground, an easily recognized one is the formation of various kinds of patterned ground. The sediments most conducive to its formation are silts and clays, especially those of glacial lake beds and alluvium (Fig. 4.52). In now temperate latitudes, the patterns show up most clearly either in spring when vegetation is beginning to grow, or during periods of drought. The former cracks, now filled with porous debris, are relatively dry so that crops growing over them become stunted or discoloured.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file Menu.htm in the Imint folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

On the Menu page, select the link Anaglyphs (Chapter 4 stereopairs), which will take you to a page that repro-

duces all the stereopairs in this chapter as anaglyphs. There are additional anaglyphs of stereoscopic aerial photography in the link Additional Images: Anaglyphs, where you can practice the skills that you have mastered in Chapter 4.

Further reading

These references comprise a useful compendium of examples of geomorphological, lithological and structural interpretation from aerial photographs and satellite images in the reflected region. Many of the selected references for Chapter 9 are also useful in geological interpretation.

- Abrams, M.J., Rothery, D.A. & Pontual, A. (1988) Mapping in the Oman ophiolite using enhanced Landsat Thematic Mapper images. *Tectonophysics* **151**, 387–401.
- Al Khatieb, S.O. & Norman, J.W. (1982) A possibly extensive crustal failure system of economic interest. *Journal of Petroleum Geology* **4**, 319–327.
- Allum, J.A.E. (1966) *Photogeology and Regional Mapping*. Pergamon, London.
- Allum, J.A.E. (1980) Photogeology—early days. *Geoscience Canada* **7**, 155–158.
- Avery, T.E. & Berlin, G.L. (1985) *Interpretation of Aerial Photographs*, 4th edn. Burgess, Minneapolis, Minnesota.
- Babcock, E.A. (1971) Detection of active faulting using oblique infrared aerial photography in the Imperial Valley, California. *Geological Society of America Bulletin* **82**, 3189–3196.
- Bailey, G.B. & Anderson, P.D. (1982) Applications of Landsat imagery to problems of petroleum exploration in Qaidam Basin, China. *Bulletin, American Association of Petroleum Geologists* **66**, 1348–1354.
- Baker, M.C.W. (1981) The nature and distribution of Upper Cenozoic ignimbrite centres in the Central Andes. *Journal of Volcanology and Geothermal Research* **11**, 293–315.
- Bodechtel, M., Kley, M. & Munzer, U. (1985) Tectonic analysis of typical fold structures in the Zagros mountains, Iran, by the application of quantitative photogrammetric methods on Metric Camera data. *Proceedings, DFVLR–ESA Workshop, Oberpfaffenhofen*, pp. 193–197. ESA SP-209, European Space Agency, Paris.
- Crosta, A.P. & Moore, J.McM. (1989) Geological mapping using Landsat Thematic Mapper imagery in Almeria Province, south-east Spain. *International Journal of Remote Sensing* **10**, 505–514.
- Dijkers, A.J. (1977) Sketch of a possible lineament pattern in North-west Europe. *Geologie en Mijnbouw* **56**, 275–285.
- Drury, S.A. (1982) A regional tectonic study of the Archaean Chitradurga greenstone belt, Karnataka, based on Landsat interpretation. *Journal of the Geological Society of India* **24**, 167–184.
- Drury, S.A. (1983) A Proterozoic intracratonic basin, thermal evolution and dyke swarms in South India. *Journal of the Geological Society of India* **25**, 437–444.
- Drury, S.A. (1990) SPOT image data as an aid to structural mapping in the southern Aravalli Hills of Rajasthan, India. *Geological Magazine* **127**, 195–207.

- ury, S.A. & Berhe, S.M. (1993) Accretion tectonics in northern Eritrea revealed by remotely sensed imagery. *Geological Magazine* **130**, 177-190.
- ury, S.A. & Holt, R.W. (1980) The tectonic framework of the South Indian craton: a reconnaissance involving Landsat imagery. *Tectonophysics* **65**, T1-T15.
- ster, N.H. & Beaumont, E.A. (eds) (1992) *Photogeology and Photogeomorphology*. Treatise of Petroleum Geology, Reprint Series no. 18. American Association of Petroleum Geologists, Tulsa, Oklahoma.
- oward, A.D. (1967) Drainage analysis in geological interpretation: a summation. *Bulletin, American Association of Petroleum Geologists* **51**, 2246-2259.
- ronberg, P. (1984) *Photogeologie*. Ferdinand Enke, Stuttgart.
- itham, E.H. (1972) Nimbus 4 view of the major structural features of Alaska. *Science* **175**, 1423-1427.
- larrs, R.W. (1974) Interpretative techniques in remote sensing. *Contributions to Geology* **12**, 23-32.
- larrs, R.W. & Raines, G.L. (1984) Tectonic framework of Powder River Basin, Wyoming and Montana, interpreted from Landsat imagery. *Bulletin, American Association of Petroleum Geologists* **68**, 1718-1731.
- lekel, J.F.M. (1978) The use of aerial photographs and other images in geological mapping. Volumes 8i and 8ii. In: *ITC Textbook of Photo-Interpretation*. International Institute for Aerial Survey and Earth Sciences (ITC), Enschede.
- iller, V.C. (1961) *Photogeology*. McGraw-Hill, New York.
- loore, J. McM. (1979) Tectonics of the Najd transcurrent fault system, Saudi Arabia. *Journal of the Geological Society of London* **136**, 441-454.
- Leary, D.W., Friedman, J.D. & Pohn, H.A. (1976) Lineament, linear and lineation: some proposed new standards for old terms. *Geological Society of America Bulletin* **87**, 1463-1469.
- andey, S.N. (1987) *Principles and Applications of Photogeology*. Eastern Wiley, New Delhi.
- ari, M.Y.H.T. (1991) Application of Landsat TM data to geological studies, Al-Khabt area, southern Arabian shield. *Photogrammetric Engineering and Remote Sensing* **57**, 421-429.
- ay, R.G. (1960) Aerial photographs in geological interpretation and mapping. *U.S. Geological Survey Professional Paper* **373**.
- icci, M. (1982) Dip determination in photogeology. *Photogrammetric Engineering and Remote Sensing* **48**, 407-414.
- ery, D.A. & Drury, S.A. (1984) The tectonics of the Tibetan Plateau. *Tectonics* **3**, 19-26.
- rv, D.A. & Lefebvre, K.H. (1985) The causes of age dependent changes in the spectral response of lavas, Craters of the Moon, Idaho, USA. *International Journal of Remote Sensing* **6**, 1433-1489.
- n, L.C. & Wetlaufer, P.H. (1981) Relation between regional tectonism and structural zones in Nevada. *Bulletin, American Association of Petroleum Geologists* **65**, 1414-1432.
- t, N.M. & Blair, R.W., Jr (1986) *Geomorphology from Space*. NASA SP-486, US Government Printing Office, Washington, DC.
- a, J.T. & Anson, A. (eds) (1968) *Manual of Color Aerial Photography*. American Society of Photogrammetry Falls Church, Virginia.
- n, M., Arvidson, R.E., Sturchio, N.C. & Guinness, E.A. (1987) Geologic mapping in arid regions with Landsat Thematic Mapper data: Meatiq dome, Egypt. *Geological Society of America Bulletin* **99**, 748-762.
- n, M., Arvidson, R.E., Duncan, I.J., Stern, R.J. & El Jaiouby, B. (1988) Extension of the Najd Shear System from Saudi Arabia to the central Eastern Desert of Egypt based on integrated field and Landsat observations. *Tectonics* **7**, 91-1306.
- n, M., Becker, R., Arvidson, R.E., Shore, P., Stern, R.J., Alf, Z. & Guinness, E.A. (1992) Nature of the Red Sea: a controversy revisited. *Geology* **20**, 593-596.
- ey, I.J. (1988) The reconstruction of palaeodrainage and regional geological structures in Australia's Canning and Officer Basins using NOAA-AVHRR satellite imagery. *Earth Science Reviews* **25**, 409-425.
- katakrishnan, R. & Dotiwala, F.E. (1987) The Cuddapah Schist: a tectonic model for the Cuddapah Basin, India, based on Landsat image interpretation. *Tectonophysics* **136**, 17-253.
- Bandat, H.F. (1962) *Aerogeology*. Gulf Publishing Company, Houston.
- dai, T. (1983) Lop Nur (China) studied from Landsat and R-A imagery. *ITC Journal* **4**, 253-257.
- nitz, E.R. (1932) Drainage patterns and their significance. *Journal of Geology* **40**, 498-521.

5 *Digital Image Processing*

Increasingly, remotely sensed images are in digital form (Chapter 3) as two-dimensional arrays made up of pixels. Each pixel is assigned a digital number (DN) that represents the energy or radiance of the waveband being monitored. Digital image handling converts these arrays of DN into signals that modulate the beam of electrons involved in producing a video display. The beam sweeps a phosphorescent screen repeatedly, so refreshing the display continuously. Images generated by this means are called rasters, so such data are commonly said to be in raster format. Desk-top scanners reproduce photographs in digital form, including the additive red, green and blue components of colour photographs. Chapter 8 explains how software convert data that do not relate to radiation, such as topographic elevation, gravitational and magnetic potential, and geochemical data that are recorded at isolated points or semicontinuously along survey lines, into raster format for display and various kinds of image processing.

The DN generally are coded in the 8-bit (1 byte) binary range corresponding to 0–255, where each step corresponds to a range in the original real data. Such byte-format data are ideal for computer manipulation using image-processing software, although modern computers are designed to handle 64- or 128-bit floating point numbers, and do so with sufficient efficiency for an 8-bit format no longer to be necessary. The degradation is for economy of storage and transmission, particularly on orbiting platforms where data rates, power and storage media are limited. It may seem that the degradation from real information presents a limit to image analysis. So it does, especially for conversion of radiance data to measures of reflectance. Moreover, the devices are tuned for all likely radiances upwelling from the surface, from the most reflective to the most absorbent. That generally means that data for any one scene are compressed to only a fraction of the available 0–255 range, rather than being optimal for the radiant properties of each scene. Where quantitative information is essential, as in sensors aimed at atmospheric composition or temperatures, such as those on the NOAA AVHRR, analogue to digital conversion involves a greater range of DN. For qualitative interpretation and many semiquantitative applications, however, the data preserve sufficient precision for them to be useful in practice. A colour image comprising RGB channels that span a real range of 6 or 7 bits (64 or 128 DN) will show up to 2^{21} or 2 million colours. That is close to the limit of human colour discrimination, and sufficient for various kinds of automated classification.

The design of digital imaging devices ensures that data from different parts of the spectrum are in perfect spatial register with one another. Thus any pixel in a scene can be thought of as a stack of DN, each layer corresponding to one waveband. Data from different devices or processes can be registered to one another, particularly to a common cartographic base using geometric correction and rectification software (Appendix B), thereby increasing the depth of the 'stack'. This registration has two important advantages. First, a colour video monitor displays three 'layers', or mathematical derivatives from them, together as additive (RGB) colour images. This is possible because video pictures are themselves made up from exactly positioned pixels of red, green and blue phosphors on the screen. The phosphors glow in proportion to the electron flux from red, green and blue colour guns, as the raster repeatedly sweeps from side to side and top to bottom to make up the picture. The DN in the image data modulate each electron flux on a pixel-by-pixel basis. The exact registration of pixels also enables the DN for different data 'layers' to be compared and combined in many different ways without disrupting the basic structure of the image.

The objective of this chapter is twofold: to introduce the methods of digital image processing, and to show how they can be applied to geological remote sensing. Some image-processing tasks aim at removing distortions and blemishes that result from imperfect means of gathering data. They are introduced in Appendix B. Those covered here relate to enhancing the quality of images for improved visual interpretation, to extracting information that cannot be perceived and displaying it as an image, and to measuring and defining degrees of statistical similarity within a scene. This last topic involves using the computer as a trained but unbiased means of recognizing spatial and spectral patterns, or classes, within the image data.

As far as possible, the treatment is descriptive rather than mathematical. There are many different image-processing software packages, and several approaches to the tasks that they perform. The account therefore is general and does not venture deeply into computer hardware and software architectures. Any geologist using an image processing package will have to become familiar with its individual foibles, its internal algorithms and the way in which it interacts with the user. This chapter provides a basis for understanding how image processing works, a guide to selecting appropriate means of processing images for various purposes, and some examples of their application.

5.1 The image histogram

An image of a single band of digital remote-sensing data is a representation of how the radiant energy reflected or emitted by the surface is distributed in two spatial dimensions. The energy is expressed as DN, which in a display are represented by a variation in brightness showing as different grey levels. Since the eye is capable of distinguishing only about 30 grey levels in a black and white image (Section 2.3), a display of up to 256 grey levels appears to be continuous. In fact it is made up of discrete steps. An image also can be expressed statistically as the probability of finding DN of a given value within it. This measure is properly termed the probability density function (PDF). The PDF is represented most conveniently by a histogram of the number of pixels, which, regardless of spatial position within an image, have a particular DN (Fig. 5.1).

The histogram of DN frequencies is probably the single most useful measure in digital image processing. Its shape indicates the contrast and homogeneity of the scene. For instance, a scene of a homogeneous surface with a low contrast will produce a histogram with a single, sharp peak. A broad single peak suggests homogeneity but a wide range of contrast. Images that contain several distinct types of surface cover may show multiple peaks. If each has a significantly different average brightness there will be a clear separation between these peaks. As their average brightness becomes more alike, so the peaks will begin to merge. Other shape attributes of the histogram give a kind of statistical shorthand for an image. The presence of 'tails' and the degree of asymmetry of the peak both indicate important structural features in an image that will rarely be obvious from the picture itself. Figure 5.2 shows examples of images and their associated histograms.

The PDF also may be represented as a cumulative histogram, with the cumulative frequency of pixels with a particular DN plotted against DN. For each DN the cumulative frequency is that of the DN plus the frequencies of all DN less than it. So the cumulative frequency at a DN of 128 is $f_{128} + f_{127} + f_{126} + \dots + f_0$. In this case the steepness of the curve is a measure of the contrast, and heterogeneities show up as kinks in the curve (Fig. 5.1).

It is of course possible to produce histograms for different parts of an image, so that the data structure in various types of terrain can be compared. These, together with the histogram of all the data, form the basic requirements for contrast stretching (Section 5.2), because it is effectively the histogram that is manipulated in this operation. A crude idea of the contribution of known types of surface to the histogram can be obtained simply by identifying known pixels and dumping their DN in different bands to a line printer or the computer console.

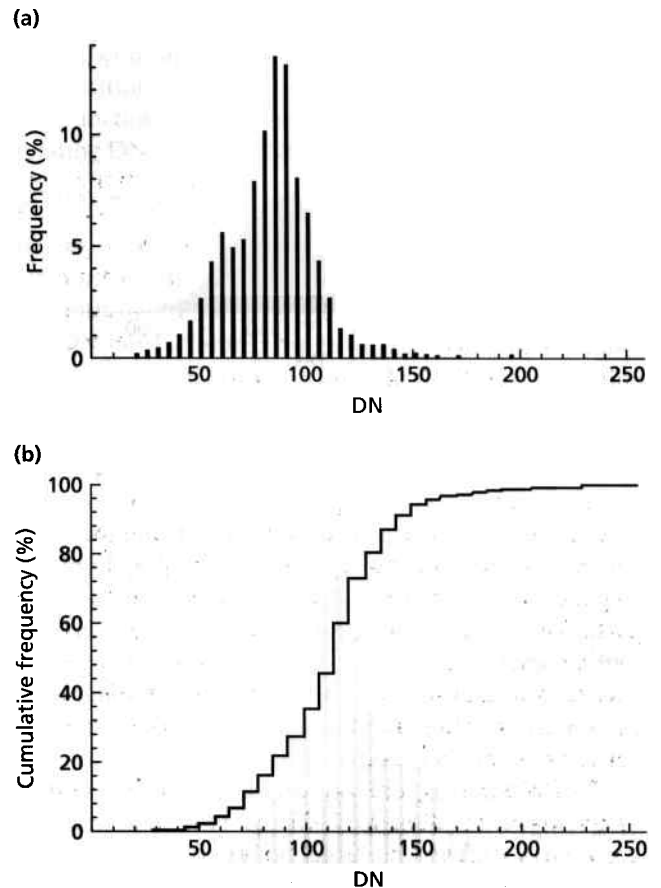


Fig. 5.1 The probability density function (PDF) of an image can be expressed by two forms of histogram. In (a) the number or frequency of pixels assigned to each digital number (DN) are plotted as bars against DN. Each division on the DN axis is termed a **bin** and is the width of one of the bars. In (b) the vertical axis is the cumulative frequency of pixels having DN less than a particular value. For example, the cumulative frequency for DN = 50 is $f_{50} + f_{49} + f_{48} + \dots + f_0$, so that the cumulative frequency for the maximum DN is 100%. The same distribution of DN appears as peaks and sometimes troughs in (a) and as changes of gradient in a roughly S-shaped curve in (b).

Multispectral data can be expressed in multi-dimensional histograms, the simplest of which is the two-dimensional histogram (Fig. 5.3). These are simply bivariate plots of DN for one band against those for another. They display visually the degree of correlation between the two bands and identify instances where a surface has distinctly different kinds of response to the two bands. The density of plots of individual pixels is a measure of where peaks in the one-dimensional histograms coincide. Highly correlated data, or classes of surface materials that have similar appearances on images of the two bands plot near straight lines on the two-dimensional histogram. Poorly correlated data show as shapeless clouds of points, as in Fig. 5.3, in which the overall poor correlation stems from the scene being a

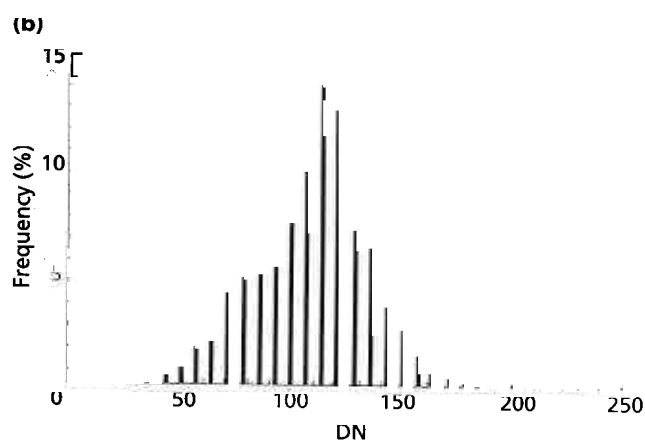
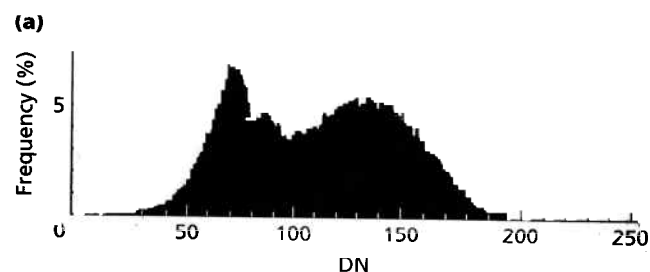


Fig. 5.2 (above) Different kinds of terrain sometimes produce contrasted histograms. Both images are of Landsat MSS band 7. In (a) the bright areas of salt flats and dark areas of lavas in the Andes are more common than surfaces with intermediate digital numbers (DN). The histogram has two peaks and is said to be bimodal. Image (b), also from the Andes, has a simple probability density function (PDF) which gives a nearly symmetrical, unimodal histogram, even though areas with distinctly different DN are visible.

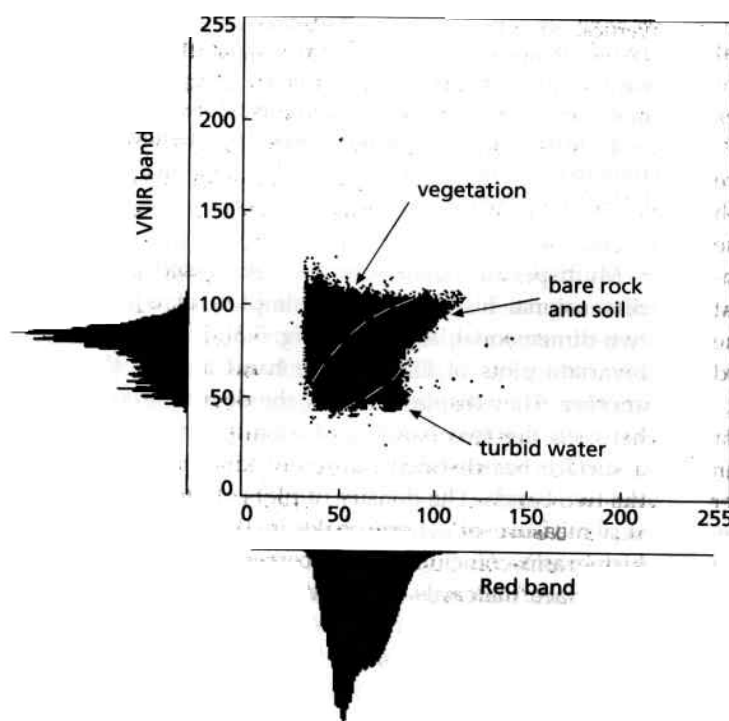


Fig. 5.3 (left) This bivariate plot of digital numbers (DN) in Landsat TM band 4 against those in band 3 is for an area comprising a mixture of open water, bare soil and rock, and variable vegetation cover. Because chlorophyll absorbs red (band 3) energy and plant cells reflect infrared (band 4) very strongly, vegetation-dominated pixels have higher DN in band 4 than they do in band 3. Vegetated pixels form a cluster which deviates from the roughly linear distribution of pixels containing rock and soil.

mixture or variably vegetated surfaces, bare rock and soil and water. The rock and soil show high correlation, and that occurs throughout scenes of arid terrains.

Histograms of multidimensional data are useful individually or combined together in deciding on the contrast stretches needed to produce enhanced colour images. Their greatest importance, however, is that they can be described statistically and manipulated as the basis for image classification (Section 5.7) or multivariate analysis (Section 5.4).

5.2 Contrast stretching

The video screen of a digital image processor, or the light source in a digital film writer, expresses DN in a stepped range of intensities from 0 (black) to 255 (maximum intensity or brightness). The simplest function of image-processing software is to change any DN in an image to any of these 256 intensity levels. It can also transform one range of DN to another. This is the basis of contrast stretching. There are many reasons why such a transformation may be required, but the most important stems from the way in which digital image data are acquired.

The objective of all land-orientated remote-sensing systems is to record the full range of reflectances possible from all conceivable surface materials. In other words real differences in the darkest and the brightest surfaces ideally should be expressed as differences in DN. The dark objects should not appear totally black and the light objects should not be 'washed-out' by saturating the sensor. Very few, if any, parts of the Earth's surface express these extremes, however. As a result the histograms for most images are compressed into a relatively small part of the 0–255 range (Fig. 5.4a). This results in loss of data, although many land-orientated systems can be adjusted to compensate for unwanted bright materials, such as clouds or snow. The human eye is capable only of discriminating about 30 grey levels in the black to white range, and then only if they are adjacent and sharply bounded. An unstretched image with a histogram that is restricted to the lower half of the 0–255 range therefore appears to have very poor contrast to the human analyst (Fig. 5.4b), even though it might contain 100 different steps. Showing sensible information means expanding the compressed histogram to the full range available in the display.

The simplest means of improving the contrast in an image is to change the range of data by spreading it equally over the 0–255 range. The minimum DN is set to 0, the maximum to 255. Each histogram bin relating to each DN originally present is moved to a new position, and the new bins are equally spaced (Fig. 5.4a). This is a linear stretch. The computer accomplishes this by

using a look-up table (LUT). A LUT converts an input ($DN(i,j)_I$) to an output ($DN(i,j)_O$), where (i,j) expresses the position of a pixel in the image. For a linear stretch the function is the basic equation for a straight line relating $DN(i,j)_I$ and $DN(i,j)_O$:

$$DN(i,j)_I = m(DN(i,j)_O) + c,$$

where, m is the gradient and c the intercept on the $DN(i,j)_O$ axis. The equation can be expressed as a graph of input against output (Fig. 5.4a). Because input and output DN have a limited range (0–255) the computer does not have to spend time performing the same calculation on each of the millions of pixels in an image. It simply replaces the input value with the corresponding output value from the LUT.

A simple linear stretch of the data in Fig. 5.4(a) produces an image that at least shows some of the surface details (Fig. 5.4c). It still appears to have a rather poor contrast, however—it appears hazy. The reason for this is twofold. First, each DN in the original data has a contribution from atmospheric scattering (Section 1.3.1), the amount decreasing as wavelength increases. Part of this is the scattering of upwelling radiance, and part the way scattered solar radiation ('sky light') illuminates the surface. The second is that the maximum DN is actually for small clouds in the image that are much brighter than the brightest part of the surface. The effect of the atmosphere is seen on the histogram of raw data (Fig. 5.4a), where there are no pixels with a DN of 0. On an airless planet total shadows would be completely black, so one way of correcting for the scattered component is to identify shadowed pixels, find their DN and set all pixels with this and lower values to 0. The chosen DN is known as the cut-off. This is termed a dark-pixel correction. Care must be taken not to choose dark pixels in water, for in the visible region their DN may be the result of suspended sediment, or a shallow bottom, not atmospheric effects. Using their value for cut-off would result in blacking out areas of dark but unshadowed surface. Another method of atmospheric correction is to plot DN for a long-wavelength band—usually in the infrared, where scattering is minimal—against those for a band in the visible region, where scattering has a greater effect. A line of best fit drawn through the distribution will intercept the short wavelength axis at a DN approximating the scattered component (Fig. 5.5). This value is then used as the cut-off. A third method is by inspection of the histograms for all the bands and estimating the shift attributable to scattering in each. Whichever method is chosen, some care is needed. Valuable information can be present in shadowed areas because they are lit by the scattering effect of the sky. Correction will lose this information. Loss of information must be traded-off against the improvement in image contrast resulting from atmospheric correction.

(a)

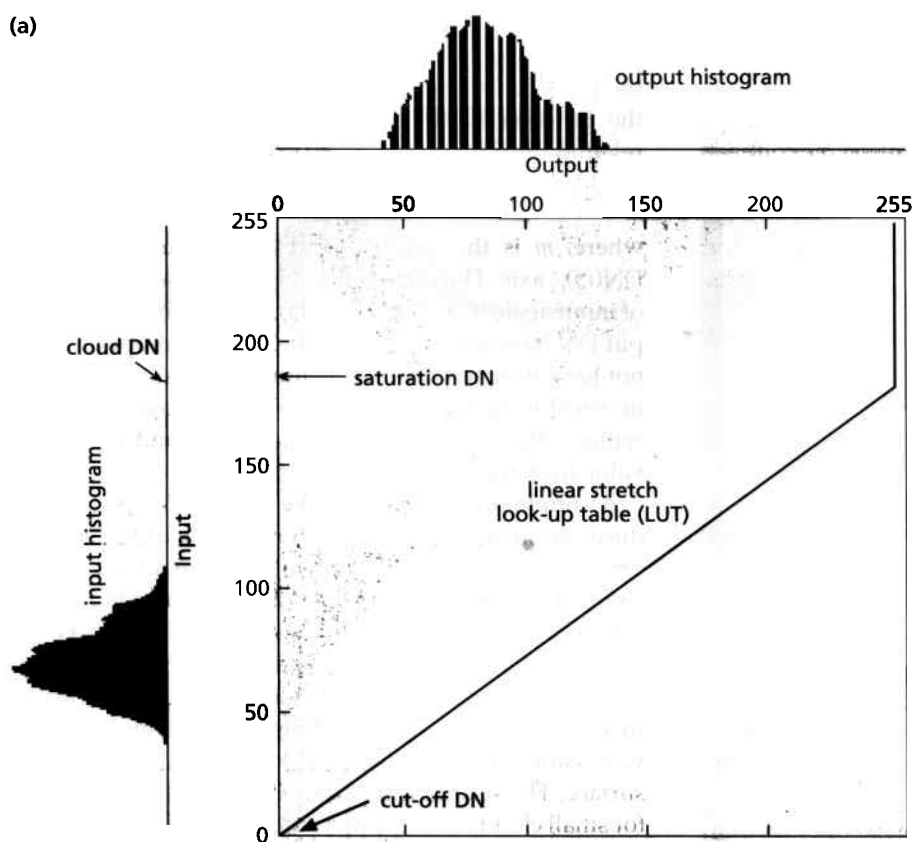


Fig. 5.4 (a) Diagram showing the histogram for raw Landsat TM band 3 data (left) for a mixed terrain on the Sudan–Eritrea border, which are compressed towards low digital numbers (DN). The simplest contrast stretch uses a linear look-up table in which DN of zero remain zero and the highest in the data (180) are set to 255, to produce an output histogram shown at the top. Display of the raw data produces a dark, low-contrast image (b) in which few details can be seen. Using the linear contrast stretch shown in (a) introduces discernible contrast in the resulting image (c). This image is approximately 30×25 km.



(b)



(c)

Figure 5.6 shows the improvement in image quality resulting from atmospheric correction of Fig. 5.4(c). As a small cloud is present in the image, and has a DN of 180 (Fig. 5.4a), the contrast has been further increased by setting the maximum DN on soil or rock to 255, thereby saturating all pixels with higher DN.

A linear contrast stretch with atmospheric correction is in most cases sufficient to produce an image of high quality, because the removal of haze and increased contrast

sharpens the high spatial-frequency features in the image (Section 5.3). Enhanced images of single bands or false-colour images comprising three contrast-stretched bands can then be interpreted geologically with a fair measure of success. They can be improved further using the interactive possibilities of the computer to suit the aesthetic preferences of the interpreter. More contrast can be introduced by increasing the value of the cut-off DN and decreasing the value of DN above which all values are

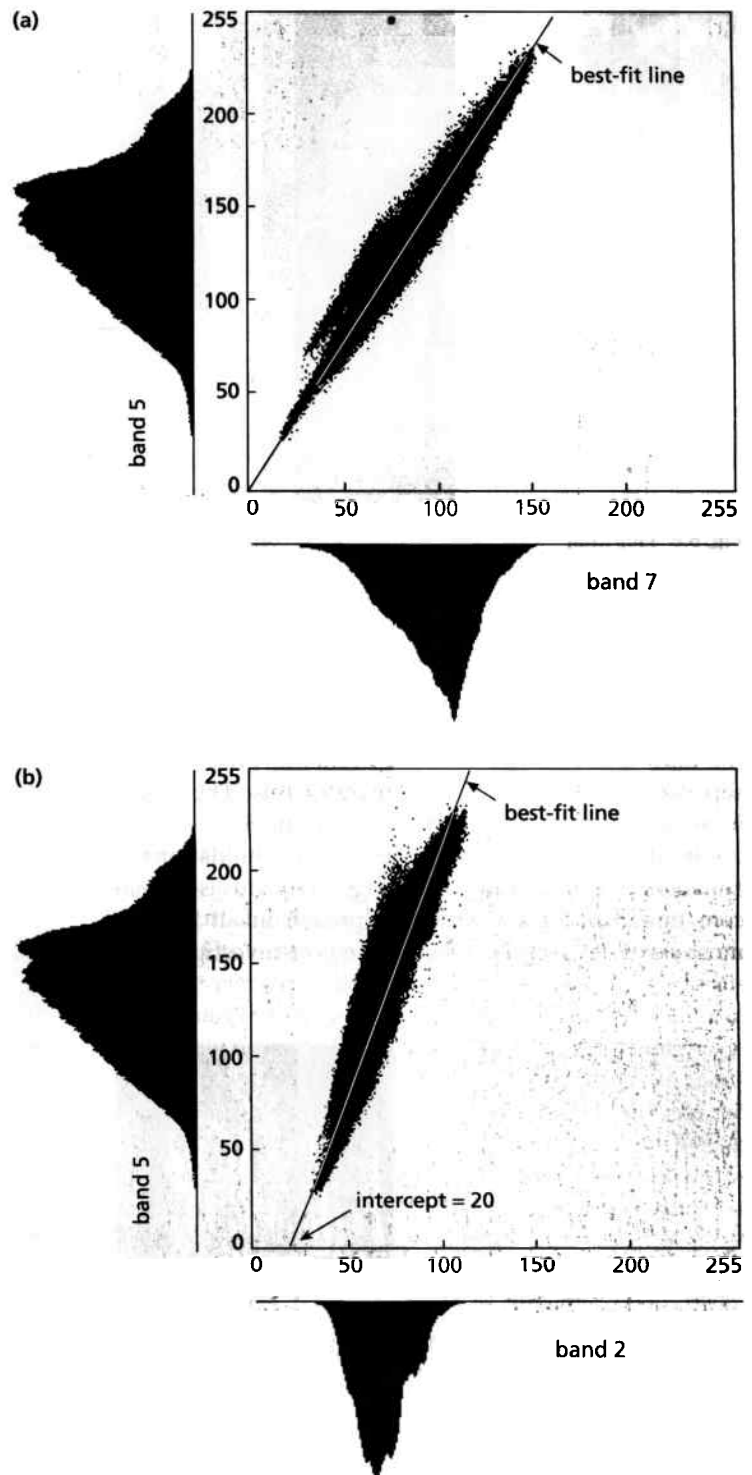


Fig. 5.5 A bivariate plot of Landsat TM band 5 against band 7 (a) has a line of best fit (the major axis of the elliptical plot), which extrapolates to the origin (0,0). This shows that atmospheric scattering is minimal in both short-wavelength infrared (SWIR) bands. A plot of band 5 against band 2 (b), which is badly affected by atmospheric scattering, gives an estimate of the scattering component in band 2 from the intercept of the distribution with the band 2 axis.

set to 255 or maximum. Increasing the cut-off will darken a larger percentage of a scene. Decreasing the DN for maximum brightness increases the image area showing as light. However, some information is lost in both cases.

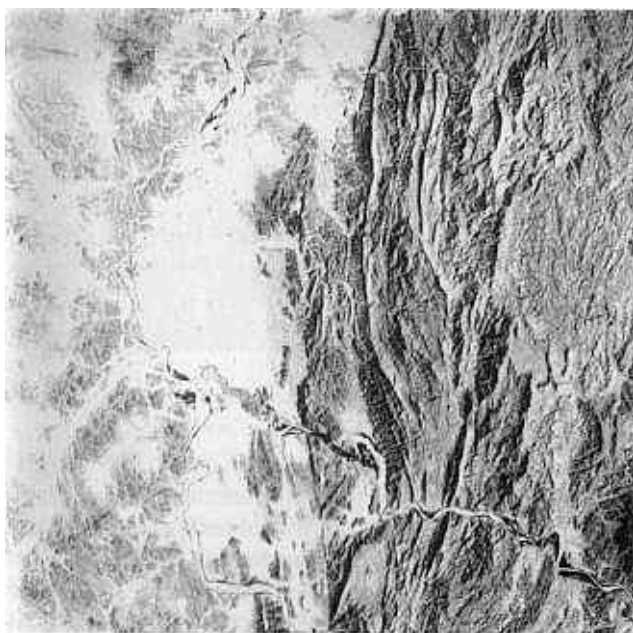
Occasionally a scene has areas dominated by both light and dark surfaces, as in the example used here, both of which contain geologically interesting information. A

simple linear contrast stretch will not perform an enhancement ideally suited to both. Contrast in the central part of the range will be enhanced at the expense of the low and high ends of the range. Again the imperfections of human vision will not allow a complete interpretation. There are many ways of overcoming these limitations by more complex methods of stretching. The simplest is



Fig. 5.6 This image is of the same data as Fig. 5.4(b), but they have been corrected for atmospheric scattering and for small bright clouds (DN = 180 on Fig. 5.4a). All the data that relate to the surface have been stretched linearly to occupy the whole 0–255 range.

to select cut-off and maximum for linear stretching of the dark and light parts of the scene, to produce two separate images. Figure 5.7 illustrates this. The image in which dark areas are stretched (Fig. 5.7a) has all the light areas washed out. The converse holds for the enhancement of the light areas (Fig. 5.7b). This is a time-consuming and inconvenient approach in all but the most extreme cases. A compromise is to investigate the

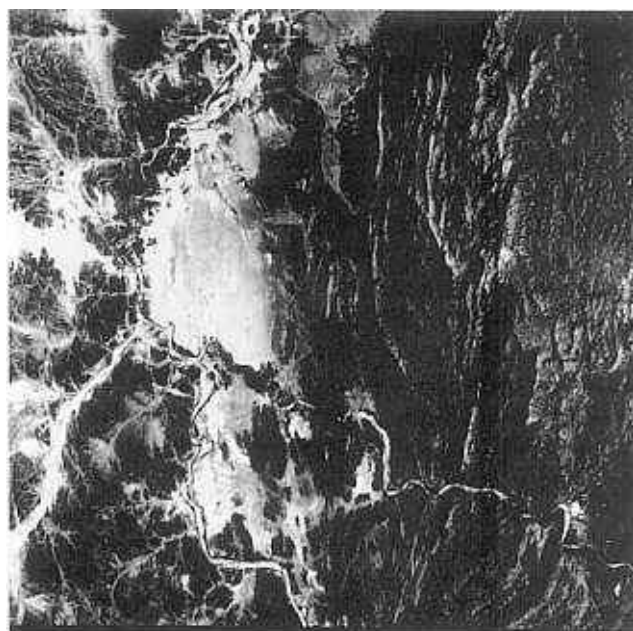


(a)

Fig. 5.7 In (a) the low digital numbers (DN) of Fig. 5.6 have been stretched linearly to occupy the full 0–255 range. The dark terrain now shows more detail, but the light areas are

histograms for the contrasted areas of geological interest and to devise a piece-wise stretch. In this operation the geologically unimportant parts of the overall histogram are compressed, and the remaining space is occupied by linear stretches of the light and dark ranges selected. Sensible contrast is thus introduced into both light and dark areas. The LUT for such a manipulation and the resulting image is shown in Fig. 5.8.

More sophisticated stretches involve creating LUTs using nonlinear transformations. To avoid the inherent loss of contrast in the tails of the original histogram in linear stretches, the raw data can be forced to resemble a normal distribution. This obeys Gaussian laws of statistical distribution, produces a bell-shaped histogram with its peak at a DN of 127 and results in a Gaussian or normalized stretch. The manipulation of bins in the raw histogram to produce an equal population density of pixels along the DN axis results in the spacing of bins becoming inversely proportional to the number of pixels that they contain; this is an equalization stretch (Fig. 5.9). This is achieved by forcing the cumulative histogram (Section 5.1) to become a straight line, and because of this the operation is sometimes known as a ramp stretch. Ramp stretches result in the greatest contrast stretch being applied to the middle, most populated range of DN, with compression in the less densely populated low and high DN tails. Sometimes it may be necessary to stretch the dark part of an image as much as possible while still retaining some contrast in the light range. This is possible using a logarithmic stretch (Fig. 5.9). The



(b)

washed out. Image (b) has had all the low DN set to 0, so that maximum contrast in the light alluvial flats allows their variation to be seen.

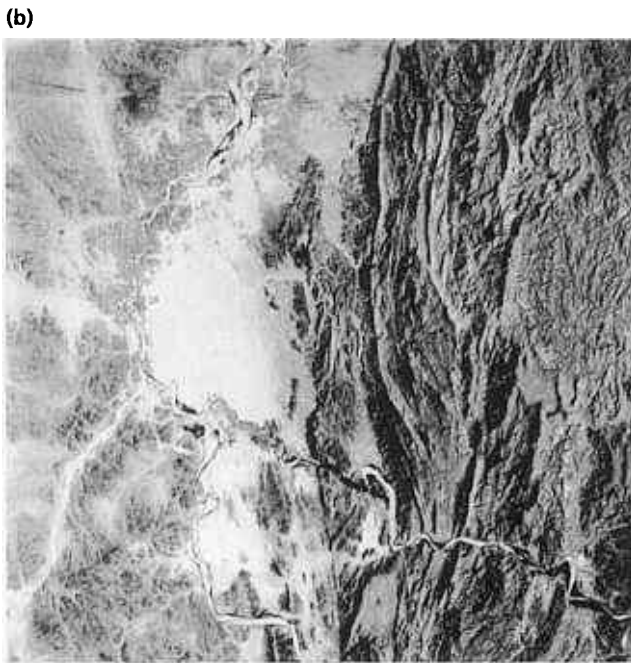
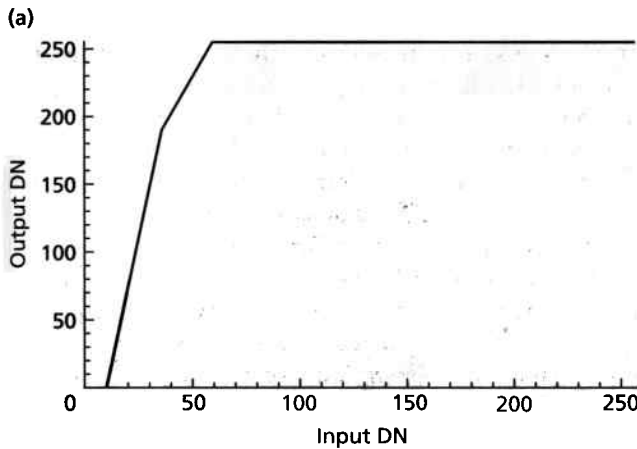


Fig. 5.8 The look-up table (LUT) (a) for a piece-wise stretch of Fig. 5.4(b) shows two different linear stretches, for the low range and the high range of geologically important digital numbers (DN). The resulting image (b) has improved contrast in both dark and light terrains. In a clearly bimodal image, such as Fig. 5.2(a), the best strategy would be to stretch the peak of low DN, compress the trough of intermediate DN and stretch the peak of high DN. In the LUT shown in (a) this would introduce a low-angle segment between the two stretched components.

converse is achieved by an exponential stretch (Fig. 5.9c). Virtually any mathematical transformation can be devised for contrast stretching, but in most cases the results are not a sufficient improvement over the simpler options to warrant the time involved.

Plate 5.1 gives examples of false-colour images in which the component bands have been interactively stretched to produce the optimum contrast and colour balance for the geological information of interest.

5.2.1 Choosing bands for colour images

The Landsat MSS and SPOT systems record data from only three regions in the visible and VNIR parts of the spectrum. Effectively, only one kind of false-colour composite image can be generated, where VNIR controls red, visible red controls green and green controls blue, to give the unfamiliar rendition of vegetation in reds and approximately natural colours for other kinds of surface. Contrast stretching can improve the range of colours in such an image, but there are limits imposed by the strong correlation between the MSS and SPOT bands (Section 5.2.2). Newer systems, such as the Landsat Thematic Mapper, JERS-1, various airborne multispectral scanners and imaging spectrometers provide more bands in the visible, SWIR and thermal infrared regions, centred on particular spectral features associated with different kinds of surface material. The data in these regions are less correlated and the number of possible false-colour images is increased. An example of just how useful this flexibility can be is shown in Plate 5.1, comparing a natural colour Landsat TM image with a standard false-colour image and one where red, green and blue (RGB) are controlled by SWIR, visible red and blue wavebands. However, the more bands that are available the greater is the choice of possible three-band combinations, including the order of their rendition in RGB. Considering the last point involves attention to the proclivities of human vision, aesthetics in particular. Quite simply there is usually a single rendition of three particular bands that is more visually stimulating (and therefore more readily interpreted) than the other five possible RGB combinations, even though the information content in all six RGB renditions is identical. Plate 5.1(c,d) illustrate this point. Although the reasons for this are not fully understood, a simple rule of thumb is to render the most informative band for a particular purpose in red, the next in green and the least informative in blue.

Choosing three bands for RGB display, when an important objective is to use as few images as possible in interpretation, is an increasingly difficult task for data with large numbers of bands. For the six reflected bands in a Landsat TM image there are 20 combinations, for each of which there are six RGB orders, compared with only six possibilities for three-band SPOT data. Clearly the potential for confusion with 120 possible colour images, let alone a variety of enhancements for each, is enormous. The numbers involved in imaging spectrometry data are greater than there are protons in the universe! Statistical analysis of the correlation between the bands can help in selecting two or three combinations for a particular objective. Inspection of Fig. 5.10, showing all six Landsat TM reflected bands, reveals that all are very similar, owing to the high correlation in the

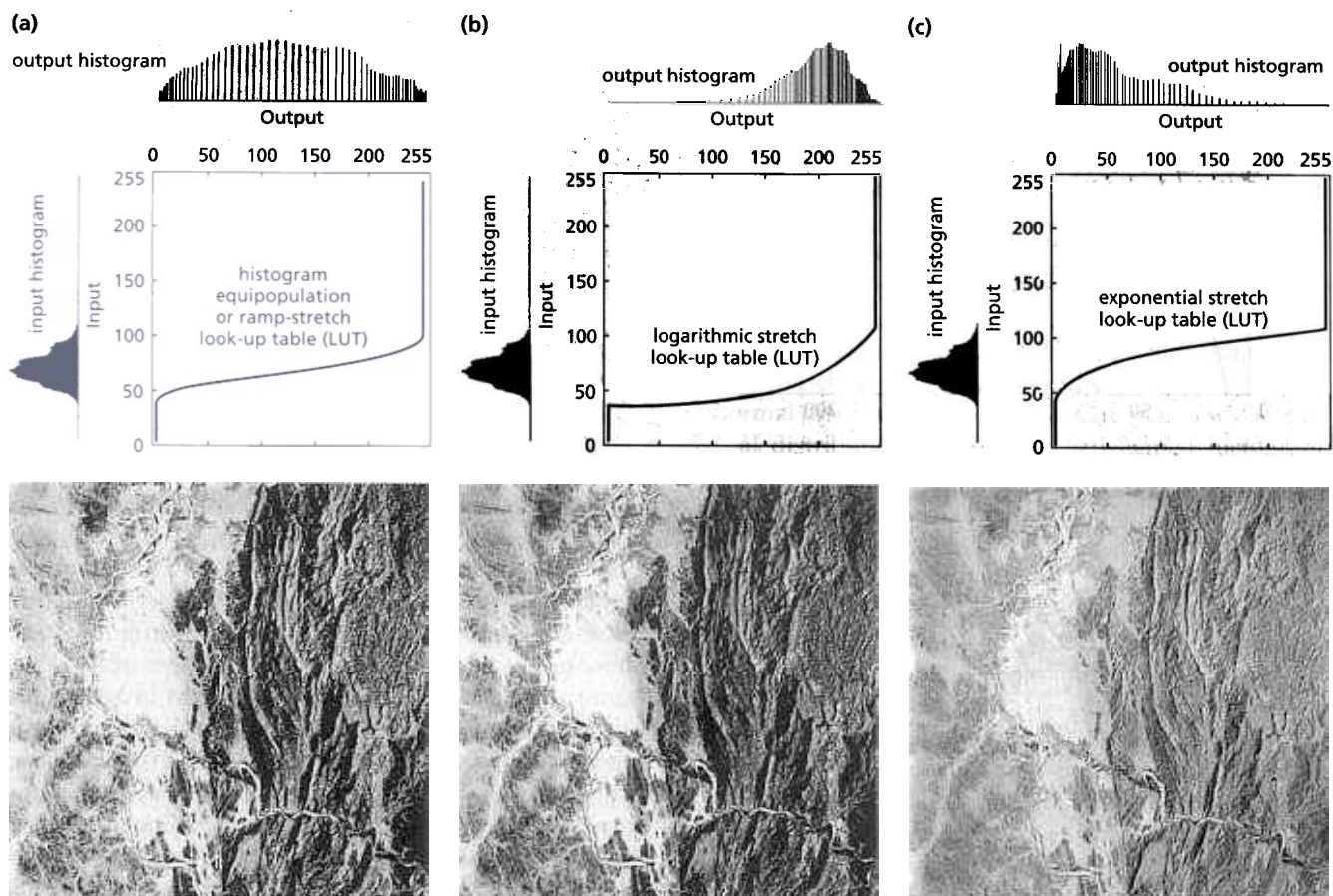


Fig. 5.9 The effects of (a) an equipopulation or ramp stretch, (b) a logarithmic stretch and (c) an exponential stretch on Fig. 5.4(b) are shown as the look-up table (LUT), input and output histograms (top), and resulting image.

reflected part of the spectrum because of the dominance of albedo. They do contain differences, but they are quite subtle. Computer analysis of the correlation can identify threefold groups of bands in order of their nonredundant information content.

However attractive, such an automatic procedure makes no distinction between the reasons for the order, which in the vast majority of cases arises from the spectrally distinct nature of vegetation (Fig. 5.3), even when it forms only a very sparse cover to an area. Although RGB combinations that emphasize vegetation differences can be useful in geological mapping, a better choice is to select bands on the basis of whether or not they contain spectral features resulting from the mineral content in soil and rock. Compared with the effects of vegetation, such features will have a very muted influence, but if the minerals involved are present their influences will be there in the data, unless vegetation cover is total. Section 5.5 discusses inspection of mineral spectra in designing uses of ratio images, but a few guidelines are appropriate here.

In the reflected region only a few groups of minerals have marked effects on band response. The dominant

ones are: Fe^{3+} oxides and hydroxides, which govern 'redness' and absorption around $0.9\ \mu\text{m}$; overall reflectivity or albedo of common rocks, which peaks around $1.6\ \mu\text{m}$; minerals that contain OH^- , which have absorptions around $2.1\text{--}2.2\ \mu\text{m}$, and carbonates with absorptions around $2.35\text{--}2.4\ \mu\text{m}$ (Section 1.3.2). The reason that Plate 5.1(c) (TM bands 5, 3 and 1 as RGB) is far more useful geologically than Plate 5.1(a) or (b) is that it expresses albedo and redness well, and the overall reflectivity and iron content of the local rocks are more variable than any other mineralogical characteristics. The narrower the spectral bands, the more strategically placed they are relative to spectral features of minerals, and the more features that are covered, the greater the possibilities for lithological discrimination.

In the thermal infrared many more minerals, including the bulk of rock-forming minerals as opposed to minor minerals and surface coatings, influence the details of emission spectra (Figs 1.12 and 1.13). Consequently, given multispectral thermal data such as TIRS or those likely to become available from ASTER, the direct mapping of true petrographic categories becomes a distinct possibility (Chapter 6).

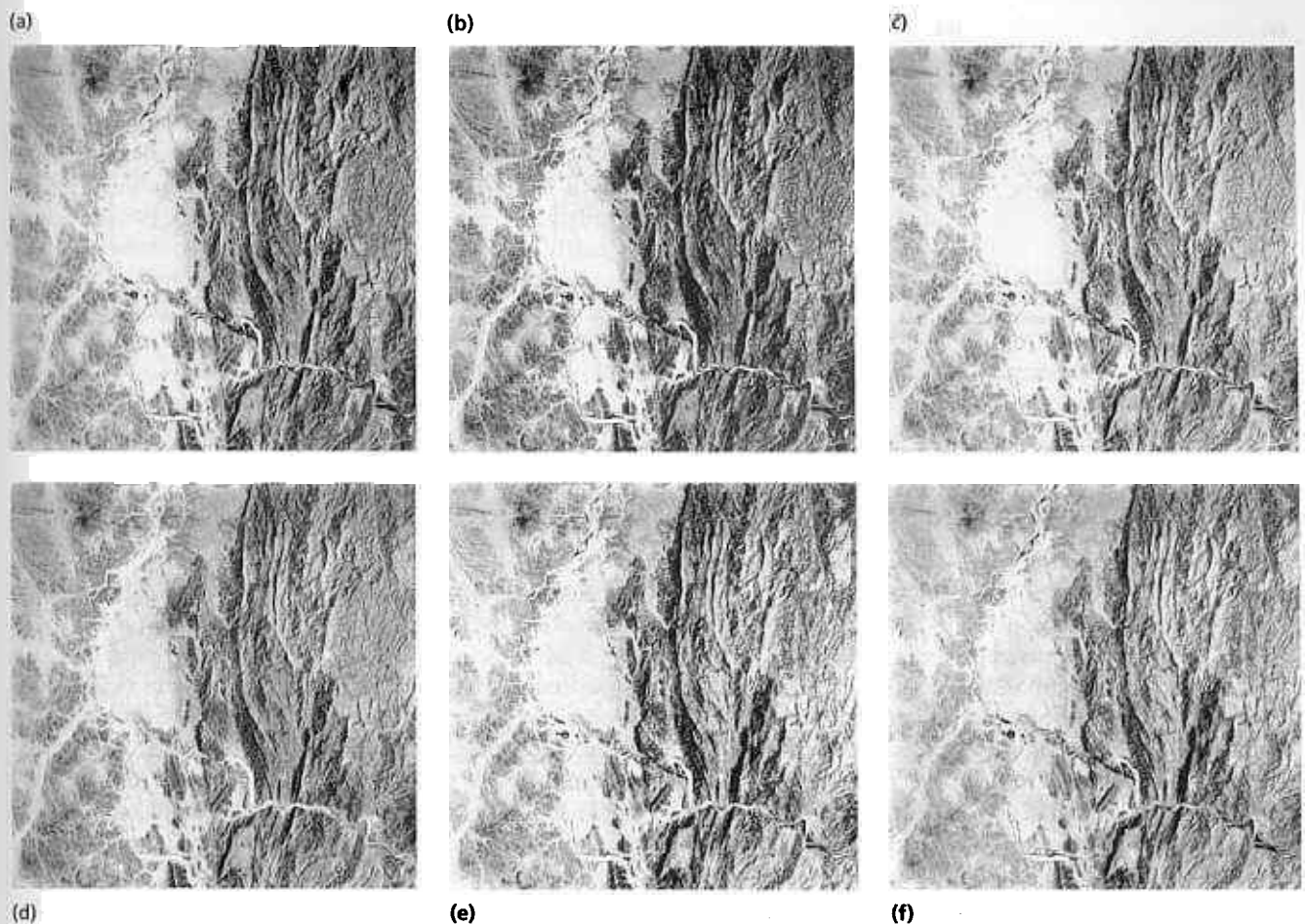


Fig. 5.10 Images (a) to (f) are of the six reflected Landsat TM bands (1, 2, 3, 4, 5 and 7, respectively) for the area at the Sudan-Eritrea border used earlier in Fig. 5.4.

In both emitted and reflected data, high correlation between bands (Fig. 5.10) is a major hindrance for two reasons. First, it means that there is a great deal of redundant data, that is, each band contains more or less the same information apart from subtleties, and for general purposes any one band could be substituted for any other. The greater the number of bands and the narrower they are, the more the redundancy. Now that is a problem for computer storage as well as a source of confusion. A seven-band Landsat TM scene occupies 250 megabytes of magnetic storage. An equivalent area of imaging spectrometry data from the HIRIS instrument would occupy about 7 gigabytes before any enhancement of it. Clearly some means of data reduction are needed (Section 5.4) to extract the wheat from the chaff. High correlation between bands is impossible to remove by simple methods of contrast stretching. The full range of 256^3 colours possible in an RGB cannot be exploited, and the bulk of colours are pastel shades that the eye finds difficult to subdivide. A means of removing the correlation must be sought for some applications.

5.2.2 Contrast stretching in different colour spaces

Figure 5.11(a-c) demonstrates schematically the way in which high correlation between two bands is retained during linear contrast stretching, and Fig. 5.11(d) expresses the limit of such stretching in RGB space. Using any of the methods of contrast enhancement discussed so far would produce the same result. The stretched data fall within an elongated ellipsoid, thereby producing only a very limited range of the colours that are possible. The full range would occupy the entire volume of the colour cube in Fig. 5.11(d). To fill that range would involve drastically reducing the degree of correlation in the data, which is the object of decorrelation stretching. This is discussed in Section 5.4, because it relies on statistical manipulation of data. The methods covered here depend on changing the co-ordinate system used to represent colour.

The major axis of the ellipsoid in Fig. 5.11(d) defines those points where the red, green and blue components are equal, thereby producing shades of grey. This is the grey or achromatic axis in RGB space, and the rest of the

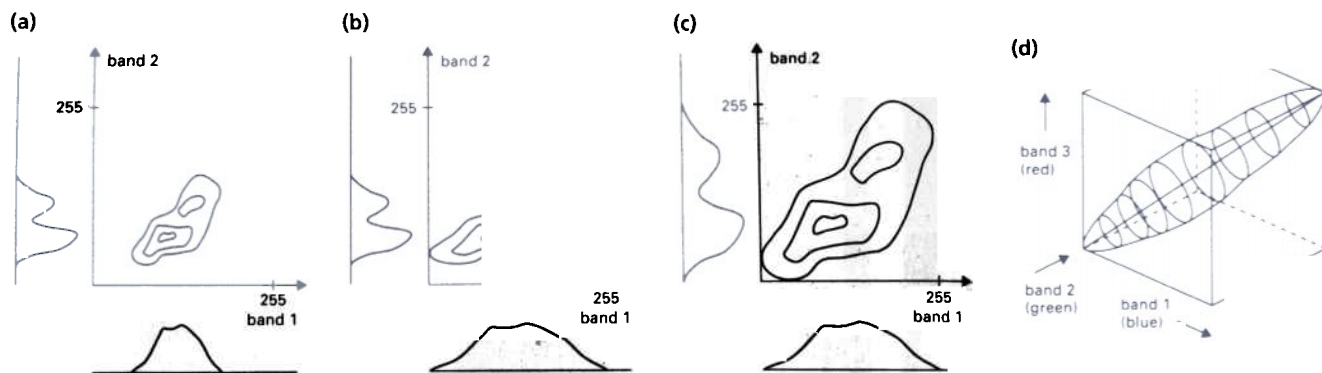


Fig. 5.11 Bivariate plots of two bands, showing (a) the raw data, (b) the data after linearly stretching band 1 only and (c) the data after stretching both bands. The high correlation between the two bands is retained. (d) Perspective view of

RGB colour space, showing the volume occupied by DN from three correlated bands of reflected data in the form of an ellipsoid with its principal axis along a line where red, green and blue DN are equal.

ellipsoid represents pastel shades and the colour limits of the associated image. Only points closer to the RGB axes represent vivid colours, and high correlation means that there can be none, except as a result of signal noise. A position on the achromatic axis, or a projection on to it from a coloured pixel is a measure of the intensity (Section 2.4), or the brightness of the colour. The actual colour can be expressed by two other measures. The closer it is to the achromatic axis the more pastel the colour, the further away the more vivid or pure is the colour. This measure is the saturation, which is measured at right angles to the axis. The third measure of colour in relation to the achromatic axis is the hue. This varies around a circle, also at a right angle to the axis. Hue is essentially a measure of the dominant colour, whether it is a pastel shade or more pure. Figure 5.12 shows the relationship between the descriptions of colour in the RGB and hue–intensity–saturation (HIS) co-ordinate systems. The HIS system is in the form of a symmetrical cone with its apex at the black point of RGB co-ordinates. Plate 5.2 is a section at right angles to the achromatic axis, showing the relationship between hue and saturation in the form of a colour wheel.

Intensity is a number from 0 at the apex of the HIS cone to an arbitrary 255 at the limit of RGB space. Saturation is an angle between 0° and half the apical angle of a cone that extends to the limits of RGB space. Hue is also an angle between 0° and 360° around the colour wheel. Red is chosen as the hue corresponding to 0° , green is at 120° , blue at 240° and the circle is completed by 360° at red again. To enable HIS space to be exploited in image processing, the RGB co-ordinates of each pixel are converted to HIS measures by a mathematical transformation. Rather than leaving saturation and hue in the form of angles they are expressed in the 0–255 range, when their ranges, and that of intensity, can be modified independently by simple stretching. Stretching intensity is akin to using the same contrast

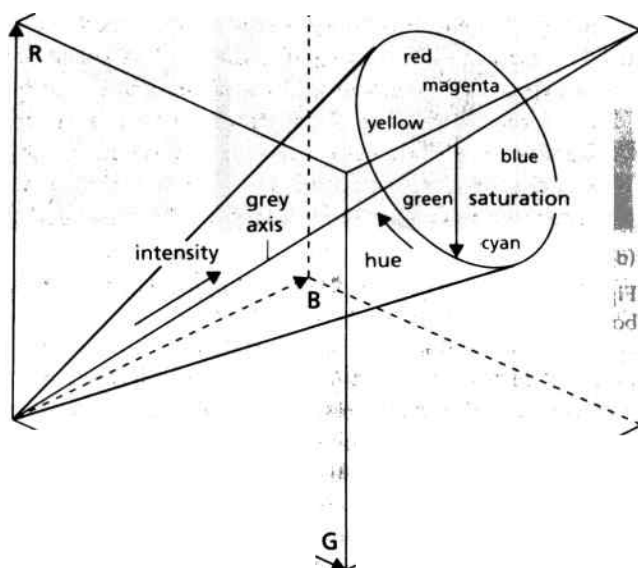


Fig. 5.12 Intensity, saturation and hue (HIS) defined in RGB space. See Plate 5.2.

stretch on the red, green and blue bands in RGB space. Because stretching hue leads to a complete modification of colour, it is normally left unchanged. Stretching saturation is the most important manipulation, for this is the same as inflating the RGB ellipsoid of highly correlated data to fill a greater proportion of available colour space. After such manipulation the modified intensity, saturation and hue measures are converted back again to RGB space by the inverse of the mathematical transform, so that the enhanced image can be displayed on a video monitor. However, most image processing software now incorporates options for direct display of HIS files as

RGB. Plate 5.3 shows the result of using a HIS transform, involving stretching intensity and saturation only, on Plate 5.1(c). The result is a far more vivid image, but one in which essentially the same colours are present — important in making a geological interpretation.

Very similar results stem from transforming RGB images into the CIE co-ordinate system (Section 2.4), but the HIS transform is simpler and takes less time. There are many variants of the HIS technique. Among them are performing other manipulations on intensity, such as spatial-frequency filtering, substituting a single band of better resolution for intensity, and combining different sorts of imagery, one as intensity another as hue. Some examples are to be found in Chapter 8.

5.3 Spatial-frequency filtering

Section 2.2 introduced the concept of spatial-frequency distribution of grey tones within a photograph. Exactly the same concept applies to digital images, with the difference that scene brightness variations are expressed in DN distributed as rectangular pixels. The variation can be expressed as a graph of DN against distance along a line on the image. The fluctuations on such a graph are the same as an assemblage of sine waves with different frequencies and amplitudes. When superimposed they interfere with one another to give the same effect as the graph (Fig. 5.13). In an image, sine waves in two dimensions express the variation in DN rather than in the single dimension shown by Fig. 5.13.

All images contain boundaries between surfaces that interact differently with EMR. They may constitute boundaries between different cover types, such as vegetation, soil types or rocks of different composition. Another kind of boundary separates areas with different amounts of solar illumination that result from topographic effects, particularly those between brightly lit and shadowed areas. In a single-band image the boundaries mark changes from one range of DN to another. In the most extreme case a boundary is an abrupt change from an area of uniform DN of one value to a uniform area with another DN. On a graph of DN with distance this is represented by a steep or even vertical gradient. Boundaries of this kind are known as edges. They occupy only a small area and are thus high-frequency features. They are narrower than they are long, and can define lines of various shapes and lengths. Good examples are the boundaries between fields with different crops, roads and rivers and their surroundings, and between shadows and well-lit slopes. Gradational boundaries, as might be expected where natural vegetation changes with altitude, have low gradients, occupy large spaces and therefore are low-frequency features. Any linear attributes they may have are vague and difficult to define.

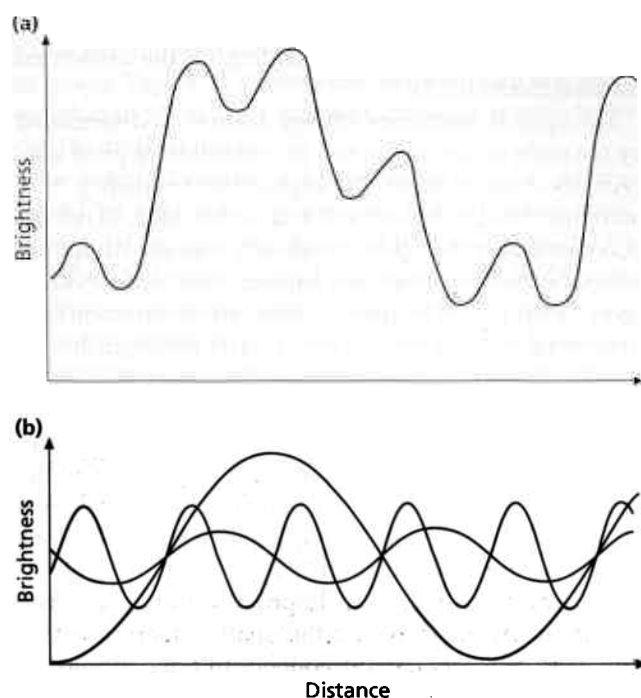


Fig. 5.13 Variations in digital number (DN) along a line on an image (a) can be expressed as a range of sine waves (b) with different amplitudes and frequencies. Their cumulative effect is the same as the real spatial frequency distribution of the image.

There is of course a complete range from high-frequency features through medium to low frequency. The scale and gradient associated with changes in DN within an image are closely interwoven, and both help in recognizing different types of boundary. For geological purposes the most important boundaries are often the edges. They may be shadow effects that give information on topographic features, some of which represent rock types with different resistances to erosion, or tectonic weaknesses such as faults and joints. Some edges are boundaries between rocks with different reflective properties, soils derived from different rock types, or vegetation boundaries that have an underlying geological control. Edges may occur in isolation, as in the case of faults or other boundaries separating large masses of uniform but different rock types. They also may be closely spaced, especially when they represent compositional banding or joints.

Whereas edges may represent small-scale geological features, medium- and low-frequency spatial features often show the gross geological features of an area. Folds may repeat stratigraphic sequences on the scale of kilometres. They define medium-frequency spatial features. Phenomena such as batholiths, major fault blocks, sedimentary basins and orogenic belts control low-frequency features with dimensions in tens or hundreds of kilometres. The lower the frequency of a feature on

an image the more subtle it tends to be, and the more difficult it is to perceive, depending on the associated contrast (Section 2.2).

The spatial frequency of any feature is determined by the scale of the image and its resolution or pixel size. With decreasing scale the higher the frequency of a particular kind of feature will tend to become. As resolution coarsens, the highest frequency features disappear altogether because they are smaller than an individual pixel. What would be undetectable subtle variation on a large-scale image often become clearly distinguished at small scale with coarse resolution. They are transformed in some cases from low-frequency features to edges. The human visual system responds to varying spatial frequencies in images in a complex way, depending on the scale and viewing distance of an image and on the eye's MTF, which is different for black and white and colour images (Section 2.2, Equation 2.1).

The process involved in improving the appearance and the interpretability of the spatial distribution of data in an image is spatial-frequency filtering. It consists of selectively enhancing the high-, medium- and low-frequency variations of DN in an image. It is not always necessary. Simply by manipulating the contrast of an image the properties of the human MTF (Section 2.2) can be exploited. Reducing the contrast leaves visible only the low- and medium-frequency features, whereas increasing it accentuates the sharp changes in contrast associated with edges. There is a limit to which either of these simple contrast transformations can be taken, however, when the visual quality of the image breaks down and interpretability begins to degrade. A more powerful technique is to use various mathematical transforms to extract or accentuate selectively the high-, medium- and low-frequency variations.

A mathematical technique for separating an image into its various spatial-frequency components is Fourier analysis, the details of which are beyond this book. The basic principle is illustrated by Fig. 5.13 for a single dimension. The Fourier transform of an image, the result of such a separation, expresses the spatial attributes of an image in terms of their frequencies, amplitudes and their orientation. It is a transform that enables certain groups of frequencies and directions to be emphasized or suppressed by algorithms known as filters. Those that emphasize high frequencies and suppress low frequencies are high-pass filters. Similarly there are medium- and low-pass filters. Moreover, selected ranges of spatial frequencies can be removed or retained in the resulting image, using band-stop and band-pass filters. The process is analogous to the electronic filtering in amplifiers to reduce hiss and rumble, enhance the bass or treble and so on in a sound recording. Filtering can be implemented through the Fourier transform, when it is said to operate in the frequency

domain, or in the spatial domain of the image itself by a process known as convolution. Frequency-domain filtering is more powerful, but is also the more expensive of computer time, involves highly complex mathematics and the result of a transform is not easily visualized in terms of the image itself. Most image processing systems routinely use convolution filters, with the option of frequency-domain filtering for special purposes.

The easiest way to understand convolution filtering is to follow the actual process that takes place in the computer itself. Convolution is accomplished by what is known as a box filter. An image can be represented by the expression or function $P(i,j)$, the original DN at every co-ordinate in the raster, where i and j are the line and pixel (row and column) co-ordinates. A box filter is a matrix of dimension $2M + 1$ rows by $2N + 1$ columns, so the number of rows and columns is odd and one cell always lies at the exact centre of the matrix. Each cell in the box filter has a weighting C , and the whole matrix can be expressed by the function $C(k,l)$, with k and l being the cell co-ordinates within the box. An output image $O(i,j)$ is computed by convolving the image $P(i,j)$ with the matrix $C(k,l)$. This operation is mathematically expressed by:

$$O(i,j) = \sum_{k=-M}^M \sum_{l=-N}^N P(i+k, j+l) C(k+M+1, l+N+1)$$

i.e., $O(i,j)$ = the sum from $k = -M$ to M of the sum from $l = -N$ to N of $P(i+k, j+l)$ multiplied by $C(k+M+1, l+N+1)$.

In familiar terms this means that the matrix C is overlain on the image with its central cell $(M+1, N+1)$ on top of a pixel (i,j) , and the other cells lying on top of the immediately surrounding pixels. The DN for each pixel overlain by the convolution matrix is multiplied by the corresponding weighting factor and the products are summed. It is this sum of the local area operations that is used to compute the value to be used in place of the DN of the pixel beneath the centre of the convolution matrix. An output image is produced by the convolution matrix systematically being moved over and transforming the DN of every pixel in the image.

The simplest convolution is that used to generate a low-pass filter. This functions by replacing the DN of each pixel with the average DN of the pixels surrounding it. Figure 5.14(a) shows a convolution matrix that could be used. In this case it is a 3×3 matrix, each cell has an identical positive weighting and the sum of the weights adds to 1.0. This ensures that the overall statistics of the image are not changed by the convolution. Figure 5.14(b-d) shows how a portion of an image is converted to a low-pass version. The dimensions of the convolution filter—sometimes called its kernel or window size—determine the extent to which the image is smoothed. Put another

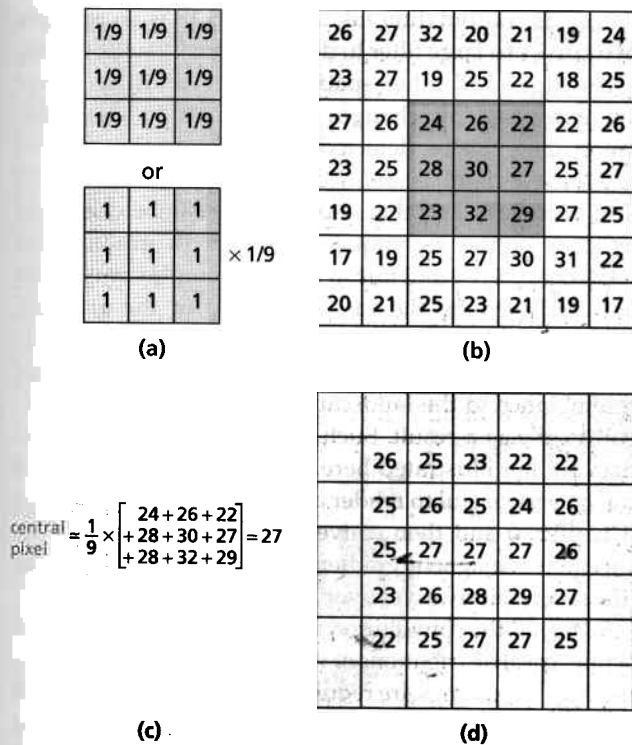


Fig. 5.14 The convolution matrix for a low-pass filter (a) comprises a 3 × 3 matrix of equally weighted cells. The data from a raw image consist of a matrix of pixels, each with its own digital number (DN) (b). During convolution, the value assigned to the pixel overlain by the central cell of the convolution matrix, shaded in (b), is the sum of the products of the cell weights and the DN of pixels that they lie over, for the whole convolution matrix. This calculation is shown in (c). The result of convolution for part of the image is shown in (d). All convolution works in this way and the results depend on the shape, dimensions and weightings of the convolution matrix.

way, the larger the filter the lower the frequency of spatial variations in DN that are expressed in the resulting image. Figure 5.15 shows the results of two low-pass convolution filters with different dimensions compared with the original image.

The simplest means of high-pass filtering is accomplished by subtracting a low-pass filtered image from the original image. This shows all the deviations in DN from the local mean. It can be achieved directly by convolution. A simple high-pass convolution filter producing differences from the local mean is given in Fig. 5.16(a). In this case the sum of the weightings is zero. The deviations from the local mean may be either positive or negative. Moreover they most often have smaller values than the original DN. So that the high-pass filtered image can be displayed, its DN must be rescaled to occupy the 0–255 range. As with low-pass filters, the size of a high-pass convolution matrix has an effect on the resulting image. In this case, the smaller the matrix the more finely spaced are the differences that show up, and vice versa. An example of a high-pass filtered image of this kind is shown in Fig. 5.17(a). One of its functions is a means of edge detection. Many other types of convolution filter can be devised for edge detection, and some of them are shown in Fig. 5.16. They may be used in cases where the simple mean difference matrix does not give satisfactory results.

A user generally hopes to perform all interpretations on a single image, using both tone and colour variations together with textural features represented by edges. The image of detected edges in Fig. 5.17(a) is noticeably lacking in contrast. Producing an image with both tonal and enhanced spatial information means adding the results of edge detection to the original image. Such edge enhancement also can be achieved by a single convolution with a suitably designed filter. Figure 5.16 shows

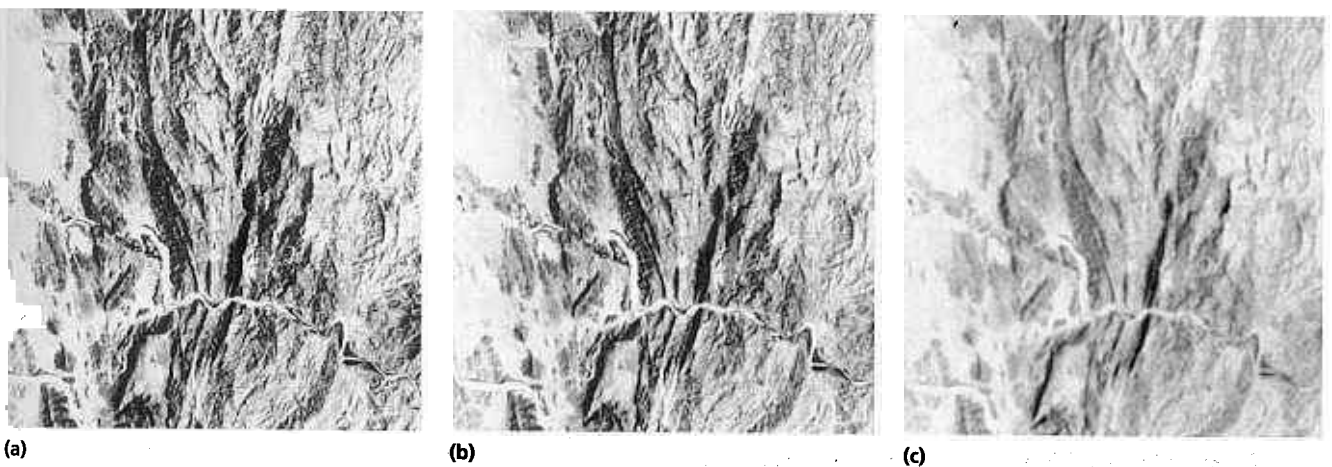


Fig. 5.15 Image (a) is a linearly stretched image of Landsat TM band 5 data for the Sudan-Eritrea area used in Fig. 5.4 and Plate 5.1. Images resulting from 3 × 3 and 13 × 13 low-pass

convolution are shown in (b) and (c). As the dimensions of a low-pass filter increase the smoothing effect becomes more obvious.

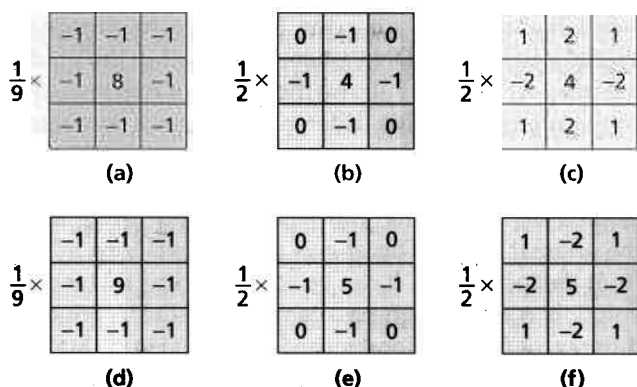


Fig. 5.16 Convolution matrices for edge detection or high-pass filtering can be achieved by subtracting the results of a low-pass convolution from the original image (a), or by devising matrices based on more complex algorithms (b and c). Edge-enhancement matrices result from adding the edge detector to the original image (d, e and f).

three edge-enhancement matrices corresponding to the three edge-detection matrices. In each case the weighting of the central cell of the matrix is increased by 1, thereby adding the DN of each pixel automatically to the convolution. The sum of the weights in each edge-enhancement matrix is thus 1. Figure 5.17(b) is an edge-enhanced version of Fig. 5.15(a) produced by this method.

A high-pass spatial filter normally enhances features that are less than half the dimensions of the convolution matrix used. In the case of geological features of regional dimensions, such as major faults and intrusive contacts, very large matrices must be used. The larger a filter, the

longer its computation takes. Figure 5.18 is an example of the use of a large filter to detect regional fault patterns.

As edges and other spatial boundaries in an image are expressed by changes in DN over a small distance, they can be detected in images of the gradient of DN. Gradient is defined as the magnitude of the first derivative and this measure also can be calculated by convolution techniques. Changes in gradient are related to edges too, but in a more complex way. They are displayed by the rate of change of the gradient, or the second derivative. Derivative techniques usually produce images of remotely sensed data that bear little resemblance to the landscape in a scene, and are difficult to use as a result. Such images of remotely sensed data are not illustrated here. In geophysics, however, it can be very useful to render aeromagnetic data as images (Chapter 8) and then convert these into images of the variation in magnetic gradient, or gradient of the gradient (the second derivative) over an area.

So far, all the convolution matrices used have involved symmetrical distributions of weightings. Sometimes asymmetrical weightings are required for specific applications. The most important of these are for enhancing spatial features arranged in different directions. Figure 5.19 shows two asymmetrical filter designs that produce approximately the value of the first derivative in two different directions. Using 3×3 matrices it is possible to achieve directional filtering to the eight principal points of the compass. Figure 5.20 shows an example of directional filtering to enhance edges running north to south across an image. In fact edges between 22.5° either side

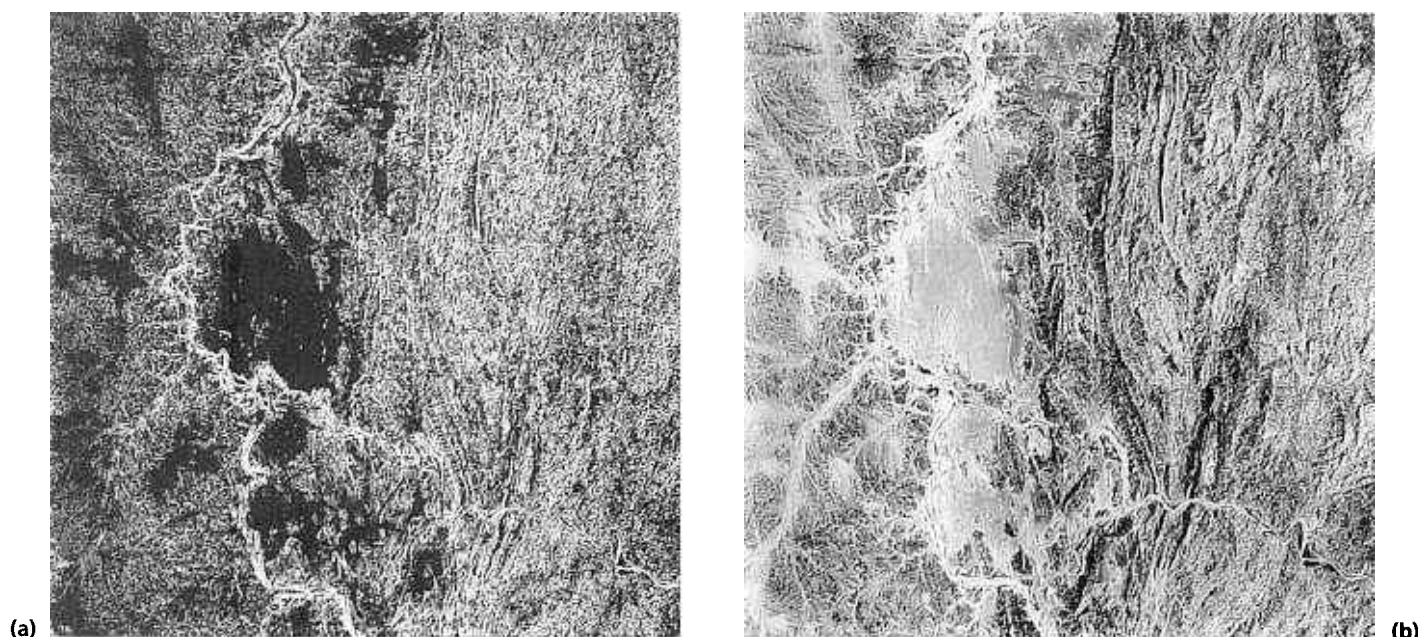
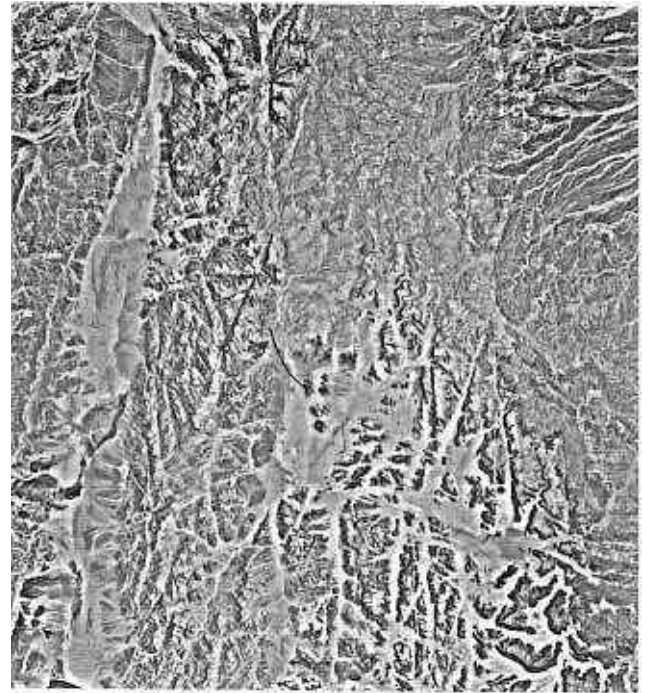


Fig. 5.17 Image (a) is the result of a simple edge-detection convolution (Fig. 5.16a) for the data displayed in Fig. 5.15(a). This was added to the original image to produce the edge-enhanced version shown in (b).



Fig. 5.18 The contrast-stretched image of Landsat MSS band 7 data for part of Jordan (a) reveals several faults that do not appear on local geological maps. However, many more are



discernible on an image of medium-scale edges detected by a 31×31 mean-difference matrix (b). Courtesy of Pat Chavez, US Geological Survey.

| | | |
|----|----|----|
| 1 | 2 | 1 |
| 0 | 0 | 0 |
| -1 | -2 | -1 |

(a)

| | | |
|---|----|----|
| 0 | -1 | -2 |
| 1 | 0 | -1 |
| 2 | 1 | 0 |

(b)

Fig. 5.19 Directional filters based on square convolution matrices depend on the symmetrical distribution of positive and negative weightings about a particular direction in the image. The results of convolution approximate to the first derivative or gradient in the direction at right angles to the axis of symmetry in the matrix. Positive gradients are normally converted to high digital number (DN) and negative gradients to low DN. Matrix (a) produces an image of gradient in a northward direction, matrix (b) in a south-westward direction.

of the chosen direction are enhanced preferentially. Figure 5.20 also demonstrates the suppression of broad tonal variations and the rendering of edges as if they were shaded topographic features. In this case this helps overcome the directional bias in satellite images as a result of illumination by the Sun. The pseudoshadowing effect of directional filters is extremely useful in enhancement of nonimage data which lack the clues to depth that actual shadows give (Chapter 8).

A danger in convolution filtering is that the filter-weighting designs are approximations only to those that would be feasible in frequency-domain filtering. Moreover, the filters are restricted to rectangular matrices.



Fig. 5.20 The image of Jordan shown in Fig. 5.18 (a) has been subjected to an eastward directional first derivative filter. The resulting image has been smoothed by a 3×3 low-pass filter and shows many faults more clearly than Fig. 5.18 (b). However, this enhancement of linear features is at the expense of tonal variations in the original image, which relate to different kinds of surface material. Courtesy of Pat Chavez, US Geological Survey.



Fig. 5.21 Directional filtering by a one-dimensional mean-difference matrix—1 pixel wide and 27 deep in this case—does highlight N-S edges in this image (see Fig. 5.17b for comparison). However, it has produced regular artefacts in two directions, roughly north-north-east and north-north-west, which could easily be misidentified as geologically significant features.

These deviations from the ideal result in spurious features or artefacts appearing in some filtered images. Figure 5.17 contains examples of a common type of artefact, known as ringing. In the raw image (Fig. 5.15a) a dark line crossing pale ground at the bottom left is an igneous dyke. On Fig. 5.17(a) this is represented by a dark line with bright fringes on either side, as are other linear features. This is ringing, which in some cases can help detect small edges, but can introduce imaginary features to the interpreter. An even more disturbing type of artefact results from the geometry of the convolution matrix, which can introduce totally new geometric features (Fig. 5.21).

Whereas using square filters avoids directional artefacts, ringing is inherent in those square filters in which all the cells have weights, the larger the filter the more gross the ringing, so that in some cases data are repeated as 'ghosts'. For small filters ringing can help accentuate small edges, but if the topography is dissected by many small features, ringing merely disrupts the patterns. Using designs with symmetrical crosses defined by zeroes and finite weights, as in Fig. 5.16(b), helps suppress ringing while enhancing edges.

5.4 Data reduction

Apart from the effects of spectral features resulting from various transitions (Section 1.3), the spectral reflectance

curves of natural materials are relatively smooth in the 0.4–2.5 μm range. In most cases a high reflectance in one waveband is matched by similarly high reflectances in the others, which is due to the result of the overall albedo of the surface. In such cases the data are said to be highly correlated. High correlation indicates that there is a high degree of redundancy among the data—for many purposes the information in one band is very much the same as that in the others. This is illustrated graphically in Fig. 5.10, which shows all six reflected bands of Landsat TM data for a scene largely bare of vegetation. A very high proportion of interband variability in scenes with abundant vegetation is the result of the unique and very broad spectral features associated with living plant cells (Section 1.3.3). High correlation between the three bands used to produce a false-colour composite results in little improvement over a single-band image. There is little colour variation. Even where bands have been selected to highlight the effects of known spectral features, redundancy can still be a major problem to the interpreter owing to the lack of colour contrast and the exploitation of only a limited region in RGB space. Very similar strong correlation between different wavebands can characterize any part of the EM spectrum where the bands are close to one another. In the thermal infrared region the correlation is even stronger than in the reflected region owing to the close approximation of many natural surfaces to blackbodies. To some extent redundancy in colour images can be overcome by manipulations such as the HIS transform (Section 5.2.2). There are, however, methods that use powerful statistical analysis to re-express the information contained in multispectral data according to its variability, so achieving a full description of the information in less variables. This is data reduction. Inherent in these methods is the possibility for inflating the data to fill more fully the multidimensional space that they occupy, thereby achieving decorrelation. This section explores some of the possibilities.

If DN from one band are plotted against those from another band near to it in the EM spectrum the majority of points lie on, or near to a diagonal line passing through the origin of the graph (Fig. 5.11a). The closer to the diagonal the data are the greater their correlation. Poorly correlated data plot away from the diagonal. The result is an elliptical distribution. Extending this to three bands often reveals the data to be mainly contained within a cigar-shaped ellipsoid (Fig. 5.11b).

A means of improving the spread of data is to redistribute them about another set of axes in multidimensional space, which maximizes the separation of differences in the data (in this case the transform is in Cartesian space, rather than the spherical system of co-ordinates used in the HIS transform). The procedure used most commonly is principal component (PC) analysis, also known as the Karhunen–Loeve transform after its originators. Means

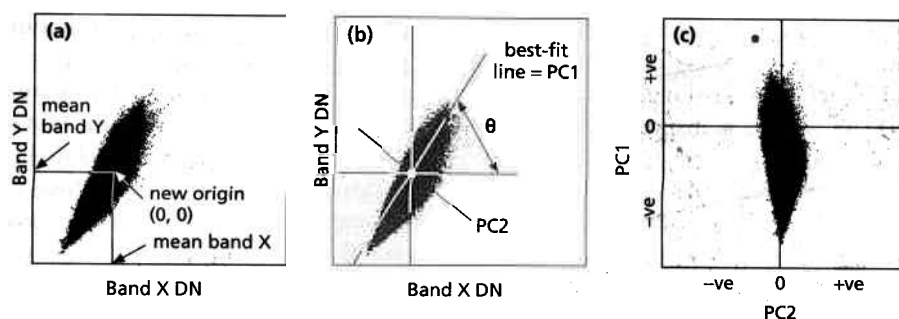


Fig. 5.22 Bivariate plots of data from any two bands in the visible and near-infrared regions generally produce an elongate ellipse of points (a) because of strong correlation. Principal component analysis begins by shifting the origin of the plot to the point defining the means (m_x, m_y) of the two sets of data (b). The axes are then rotated through an angle θ so that one is aligned with the maximum variance in the data (c).

This axis becomes the first principal component on to which new values of the data, combining contributions from both bands, are projected. The second axis expresses the variance that cannot be expressed by the first principal component. When these residual data are projected on to this axis they comprise digital numbers (DN) for the second principal component.

and variances of data from two bands show graphically as scatter in a bivariate plot (Fig. 5.22a). The variance of a single band expresses the spread of values about its mean. A measure of the joint variation of two variables is their covariance. When covariance is positive the data are positively correlated, as indicated by the crude slope towards the origin of the cloud in Fig. 5.22(a), and when it is negative an inverse relationship holds. When the covariance is zero, the two channels of data are completely independent of each other; a bivariate plot would show a formless cloud. For multispectral data (topologically speaking multidimensional) the full statistical relationship between all the axes is expressed as a covariance matrix. A matrix of correlation coefficients expresses the degree of data interdependence in another way.

The first step in PC transformation shifts the origin of the n -dimensional distribution to that point which represents the means of all the bands. This sets the means to zero—the procedure is easier to visualize in the two-dimensional case (Fig. 5.22b). This step allows rotation of the axes, until one coincides with the line along which all the data show the greatest spread, along the principal axis of the ellipse of data (Fig. 5.22c). This new, rotated axis is the first PC. Values associated with PC1 are projections of the 'raw' data points on to it at right

angles. An axis perpendicular to this now defines the line expressing all the remaining variation in two-dimensions, again by orthogonal projection to the axis from the data points. This is the second PC. In space with more than two dimensions this operation continues to define orthogonal axes that progressively consume all the variation that does not project on to lower order PCs. It results in as many PCs as there were original data axes. The mathematics to achieve this transformation uses an expression of the intervariability of the constituent 'raw' data bands, either the covariance or the correlation matrix. The results from the two alternative 'driving' statistics are slightly different, but the reasons need not be explained here (for simplicity the covariance matrix is used). In short, PC analysis is a little like finding the axes of a deformed pebble and measuring them for strain analysis, one big difference being that it deals with hyperspace.

Projection of the original n bands of data on to n new PCs uses eigenvectors to produce linear, additive combinations of the original data for each PC. Each eigenvector is derived from the 'driving' statistics, and is a weighting or loading factor for the additive contribution of each band to a PC that can be positive or negative, depending on the particular geometry of the data distribution. Table 5.1 includes a matrix of eigenvectors used to project DN from each band in the Landsat TM scene in

Table 5.1 Eigenvectors for each principal component of the Landsat TM data shown in Fig. 5.23 indicate the relative loadings of the different bands in each component. The eigenvalues indicate the relative proportion of the overall scene variance encompassed by each component

| | | Eigenvectors | | | | | |
|-----|-------------|--------------|--------|--------|--------|--------|--------|
| | Eigenvalues | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 7 |
| PC1 | 2242 | 0.27 | 0.21 | 0.37 | 0.31 | 0.68 | 0.42 |
| PC2 | 284 | 0.58 | 0.32 | 0.40 | 0.27 | -0.48 | -0.32 |
| PC3 | 18 | -0.48 | -0.11 | 0.11 | 0.67 | 0.18 | -0.52 |
| PC4 | 11 | -0.44 | 0.01 | 0.37 | 0.23 | -0.50 | 0.60 |
| PC5 | 6 | 0.31 | -0.14 | -0.67 | 0.57 | -0.16 | 0.29 |
| PC6 | 1 | -0.27 | 0.91 | 0.32 | 0.00 | 0.00 | 0.00 |

Fig. 5.10 on to each PC. The value assigned to a pixel's PC is simply the addition of the products of multiplying the DN for each band by its associated eigenvector (the columns in Table 5.1). As well as transposing the axes through the n bands of data, PC analysis transforms the covariance matrix too, so that the covariances between the resulting PCs are set to zero. In other words the PCs are decorrelated. The variation within the 'raw' data is recast as variances on the PCs, which are known as eigenvalues. The highest eigenvalue is associated with the first PC, and progressively lower eigenvalues with each higher order PC. Figure 5.22(c) show that the values assigned to a PC can be positive or negative, because of the initial repositioning of the origin in the distribution. For them to be displayed as an image means expressing the range of values in the familiar 0–255 range using a look-up table of some kind.

The first PC of highly correlated data is generally a weighted average of all the data. In the case of reflected data, this approximates an image of albedo combined

with variations in solar illumination that result from topography. In the case of multispectral thermal data, the first PC is dominated by surface temperature variations. Atmospheric effects reside in the first component because they are always highly correlated. The higher order components express deviations of various kinds from the average in PC1, and contain information relating to spectral variations in the scene, which stem from different types of surface. Because noise in the data generally is the least correlated between bands, it ends up in the higher order components. Figure 5.23 shows the six PCs derived from the data shown in Fig. 5.10. The decrease in variance and increase in noisiness as the order of the components increases is obvious. But what do these new images show, after such a complicated process?

Making sense of PC images involves examining the eigenvectors, or relative loadings of each input band to each component. As Table 5.1 shows, the loadings are all positive and high for the first component, indicating that it expresses the average of all the bands quite closely.

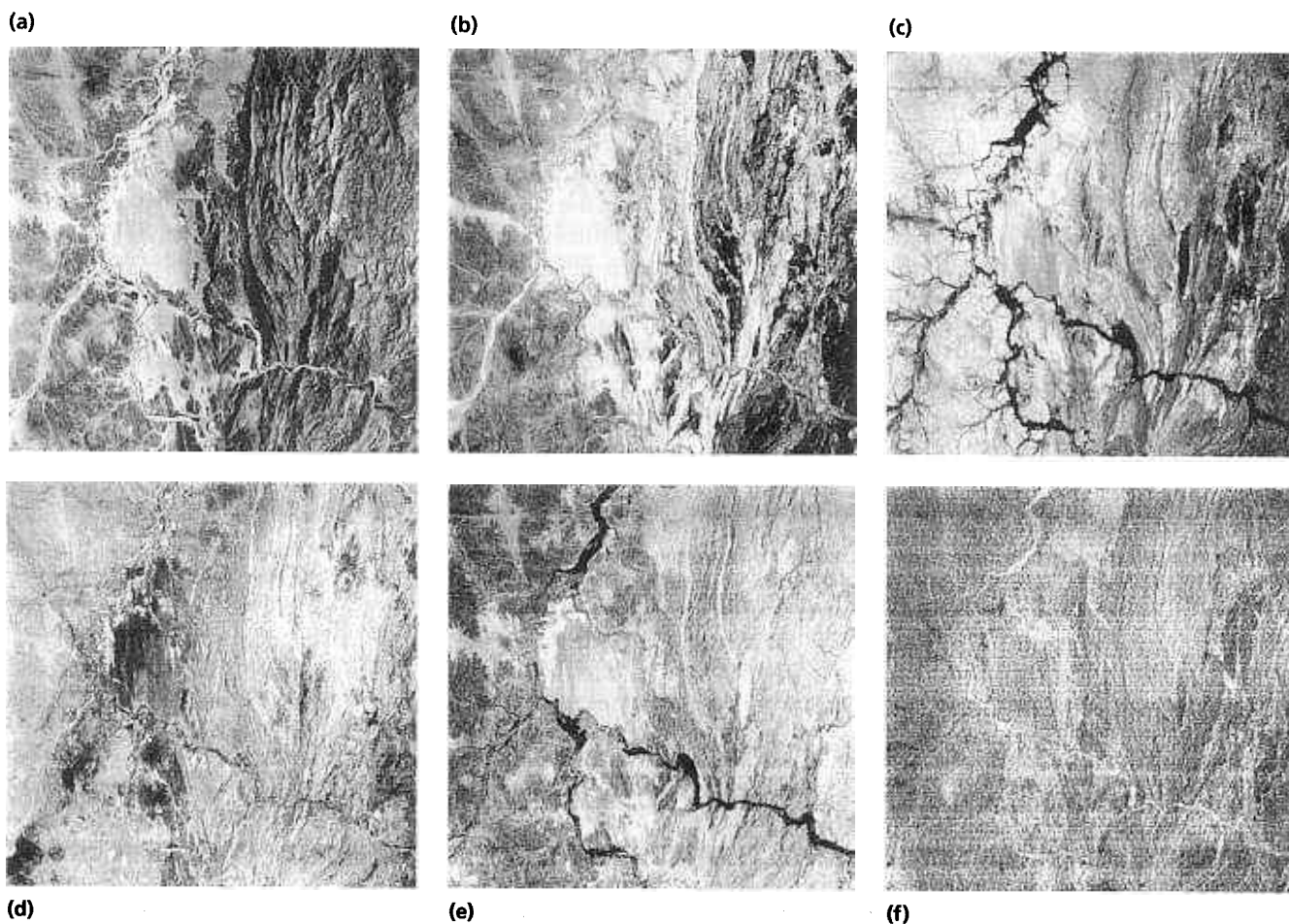


Fig. 5.23 Images (a) to (f) represent data in the six principal components that are possible for Landsat TM reflected data. They cover the same area as Fig. 5.10. The high quality of the image of the first principal component results from the expulsion of noise to higher order components, and from

combination of highly correlated contributions from all six TM bands. The bulk of noise resides in the sixth PC (f). The other PCs are additions of each band weighted by the respective eigenvectors as explained in the text. A false-colour image derived from these data is shown in Plate 5.4.

However, for the other components there is a mixture of positive and negative, and low and high eigenvectors. Understanding these numeric clues in the eigenvector matrix means returning to specific features in the spectra (Chapter 1). Considering vegetation first, this has a uniquely high reflectance in Landsat TM band 4 and a low one in band 3, relative to soils and rocks (Fig. 1.15). A component that has a high eigenvector for band 4 and a high value with the opposite sign for band 3 (e.g. PC5 in Table 5.1) is clearly expressing variations in vegetation density. Figure 5.23(e) shows dark features along drainages that relate to vegetation, which are where expected from its red appearance on the 432 image in Plate 5.1(b), but which occupy a larger area than suspected simply from a conventional false-colour image. (Note: whether materials associated spectrally with this or that PC show prominently as dark or light relative to their surroundings is impossible to predict.) Clay minerals' high reflectance in band 5 and strong absorption in band 7 (Fig. 1.9) mean similar high loadings in a component that contains information about clay variability (e.g. PC4 in Table 5.1). Figure 5.23(d) shows dark areas, but to know if the prediction is verified would mean field checks in this case (in fact most of them are clay-rich alluvium). Much the same argument for the spectral characters of ferric iron minerals in the visible and VNIR (Fig. 1.7) can help identify a component that expresses iron variability. Unfortunately things are rarely that simple, and the eigenvectors generally are a guide only to understanding PC images. There is, however, a ruse to force PC analysis to enhance specific materials selectively (Section 5.4.1).

As the effect of PC analysis is to compress the variance of the original data into the lower order components, it is a means of data reduction. This is rarely useful for data with few, broad bands, such as Landsat TM. Important information for small, spectrally unusual parts of a scene may lurk in the higher order components from such data, so it is important to check all components carefully. A PC approach is much more useful as a means of data reduction in analysing the huge volumes of hyperspectral data from imaging spectrometers (Section 5.6.2). With such a wealth of data there are many possible combinations to examine, and confusion is inevitable using simple band techniques.

Principal components may be displayed in any combination as red, green and blue to produce a false colour image. Instead of the frequently limited range of colours resulting from normal band composites, a vivid array is common (Plate 5.4). This effect of the decorrelation process can be extremely useful in separating subtly different categories of surface. However, the colours bear little relationship to the true composition and EM properties of the surface. To explain them would mean considerably more than grappling with the intricacies

of the eigenvector matrix, for no two scenes are alike and the covariance matrices for each produce unique sets of components and associated eigenvectors. So, images of different terrains and images of the same scene acquired on different dates will not show similar geological materials in the same way. Rather than attempting to identify the reasons for the appearance of specific divisions of the surface that are not seen in conventional images, it is better to use colour PC images to map those divisions and the structures that they reveal arbitrarily, and then seek reasons for their separation in the field. There are other strategies though.

5.4.1 Directed principal component analysis (PCA)

Figure 5.3 shows how different materials express themselves in bivariate plots, in this case for a plot of VNIR against red radiance for an entire Landsat MSS scene. The scatter is dominated by an ellipse showing high positive correlation that corresponds to bare soil and rock. Water shows its characteristic low VNIR and significantly higher red radiances and forms a separate cluster towards the bottom. Vegetation with strong VNIR reflectance and red absorption forms the negatively correlated 'lump' on the main cloud. Similarly, iron- and clay-rich surfaces would form identifiable clusters on plots, say, Landsat TM bands 3 and 1, and 7 and 5, respectively. In the case of the VNIR-red distribution in Fig. 5.3, the first PC axis clearly would not follow the axis of the main highly correlated ellipse, and its position in n -dimensions also would be significantly shifted in areas with a high proportion of vegetation. The outcome of this is that in such scenes PCA becomes dominated by the vegetation effect, and increasingly suppresses the separation of geologically interesting surfaces, depending on the proportion of vegetation cover.

One means of using PCA to accentuate geologically controlled spectral variation is deliberately to omit densely vegetated parts of a scene from the statistics that 'drive' the process. An arbitrary 'rule' is applied, as metaphorically in ecclesiastic law! So this approach is known as canonical analysis. In practice the operator outlines a large area with the lowest vegetation cover, generally using a VNIR, red and green rendition in an RGB image. This training set then becomes the source for the covariance matrix, and the PCA begins with the PC1 axis forced to follow the principal axis of the soil and rock pixels. Two things should happen: variance as a result of vegetation appears strongly and almost uniquely in one PC—often the second; variance as a result of rock and soil spectra becomes more accentuated in the other high-order PCs, and related more easily to spectral properties of important minerals. Theoretically, this should produce more useful results than plain, vanilla PCA. In practice it suffers from one main drawback; the

ults can never be repeated for other scenes, even those acquired on the same date—the adjacent rows along a transform's path. In addition, it becomes tempting to refine an area for the statistics, or to try other areas where, perhaps, more is known of the field geology. Time soon passes by, and with it an increasing and a confusing mass of different PCs results. Where geological knowledge is lacking, canonical analysis can be hit or miss, but it does prove useful where some areas are known to show interesting mineral associations, as in the case of hydrothermal alteration. Using them to define a training set would result in other, unknown areas with similar properties to show up in the resulting PCs. The other way of directing PCA is to focus on general knowledge of spectral properties, without making arbitrary decisions about whether this or that part of a scene is most representative of geology.

Examining the spectra in Chapter 1 shows that in some cases a feature characterizes a particular material. Vegetation is a case in point, and can be defined from the relative difference in VNIR and red bands. A ratio between the two is quite sufficient to highlight vegetation (Section 5.5). The same could be said for ferric oxides and hydroxides, as they are uniquely red-tinged and ought to show in a ratio image involving red and blue radiance. There are differences between the oxides and hydroxides, however, the latter being orange-brown and the former more reddish. Besides that, both have absorption features in the VNIR. If data for these regions are available, as they are in the cases of Landsat TM and ASTER, using all of them would be much more effective, particularly if data more suitable for other materials were excluded. The same applies to minerals containing Al–OH, Fe–OH or C–O bonds that produce absorptions in the 2–2.4 μm range (Figs 1.9 and 1.10). Using Landsat TM bands 7 and 5 data alone would lump them all in a single category because the breadth of band 7 covers all the features. However, clays (Al–OH) also show a higher ratio of blue to red radiance (they are 'cream' coloured). Carbonates are spectrally flat in the visible region (they are 'white'), whereas Mg–OH minerals, such as malachite, are often greenish. So data from both the visible and SWIR regions has the potential for at least some subtle discrimination, when broad-band Landsat TM is the only sensor available.

The general strategy adopted by most researchers (see Further Reading), who have ventured into this particular 'haunted wing' of PCA, is based on accepting that the only clear discrimination possible with Landsat TM data is between ferric oxides/hydroxides and OH- and C–O-bearing minerals. They group the available TM bands into 1, 3, 4 and 5 (for ferric minerals), and 1, 4, 5 and 7 (for OH- and C–O-bearing minerals). Principal component analysis of these groups produces four PCs with associated eigenvectors. That PC with high (positive

or negative) loadings for bands 1 and 3 in the eigenvector matrix gives a measure of 'redness' that is related to oxidized iron minerals. Similarly, an OH/C–O PC will have high loadings for bands 5 and 7 in the appropriate eigenvector matrix. Forming a composite channel, by adding the two PCs together, gives a third channel to use in an RGB image. Such images have limited hues, but by this simplification they clearly outline areas rich in iron, OH/C–O or both mineral types. The last is of considerable importance, for many zones of hydrothermal alteration show both in abundance, and often have associated metal mineralization.

Potentially, as explained earlier in this section, this type of directed PCA should permit the discrimination of several more mineral groups. The grouping of TM bands 2, 3, 4 and 5 should distinguish haematite from goethite, because of the significantly higher reflectance of the latter in the green region. Using TM bands 1, 3, 5 and 7 should give some information about the Mg–OH and C–O bond distortions as well as that due to Al–OH. Careful selection of high-order PCs from the two groups, with reference to the respective eigenvector matrices, and display as RGB images, and some knowledge of the geology covered by a scene, potentially results in highly processed images and discrimination of subtle features. With the failure of the Lewis platform with its hyperspectral sensor, and the indecision regarding the deployment of HIRIS in the EOS series, Landsat TM and ASTER will remain the main source of geologically oriented image data for some time. In the last case, directed PCA that extends into the multispectral thermal bands will well present a quantum leap in routine geological image interpretation. However, hyperspectral data will inevitably colour the future of geological remote sensing from orbit, and Section 5.6.2 focuses on that, including a note on directed PCA.

5.2 Decorrelation stretching

Figure 5.24 shows in two-band form the principle of decorrelation or D-stretching. Figure 5.24a shows two highly correlated bands relative to original band axes and new PC axes. Figure 5.11 and Section 5.2.2 explain that there are limits to the exploitation of RGB colour space by simple contrast stretching. The rotation of axes involved in PC analysis enables the cloud of data to be stretched in two directions at right angles, instead of just along the major axis of the original elliptical distribution. The space defined by the new axes can be filled more efficiently (Fig. 5.24b,c). Simply by rotating the axes back to their original position in band space (Fig. 5.24d) results in extreme, and otherwise impossible, 'inflation' of the data. Moreover, this procedure retains the relative position of the pixels in the original distribution. This means, in the three-dimensional case used in RGB images,

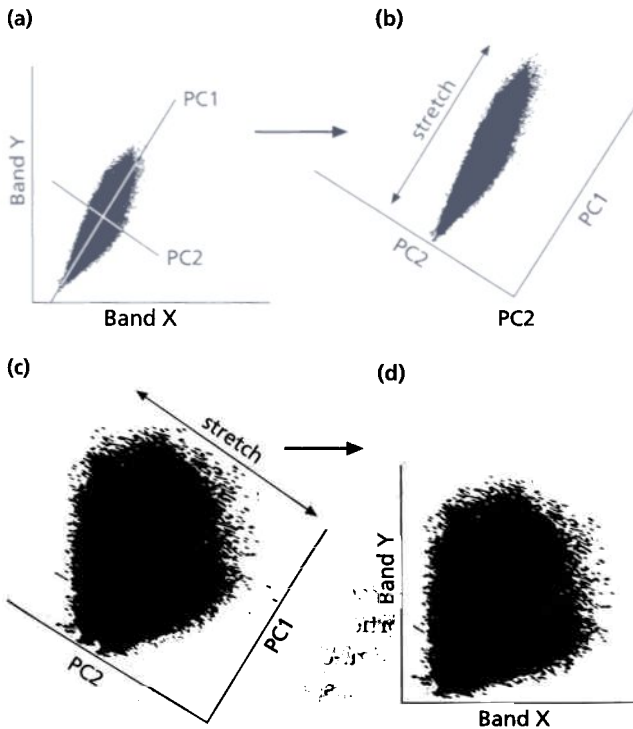


Fig. 5.24 (a) is a bivariate plot of two bands with principal component (PC) axes superimposed. In (b) the first PC has been stretched after rotation of the axes to PC space, the second PC having been stretched in (c). This shows how it is possible to produce a decorrelation of the data in PC space. (d) shows the rotation of the decorrelated data back to original band space, the original correlation having been drastically reduced, although each pixel is still in its original position relative to all the others. Compare with Fig. 5.11.

that a colour image can be stretched by this means to give nearly all conceivable colours, without altering the basic hues of each type of surface in the scene. This is achieved by stretching the PCs to the full 0–255 range and then reversing the transform. Complex as it may seem, in practice this is merely adding together the products of multiplying all the stretched PCs by their associated eigenvectors—the rows in Table 5.1—for each band to produce the new D-stretched bands. As an example, band 1 is reconstructed by adding the products of each PC and its associated eigenvector in the top row. Plate 5.5 shows a D-stretched version of Plate 5.1(c) for comparison with Plate 5.3. The effect is quite similar. Some of the unique possibilities presented by this technique, however, have been used to sharpen the image and to reduce the noise that is inevitable in extreme stretches.

As in the manipulation of intensity in the HIS stretch, using edge enhancement sharpens the first PC (Section 5.3) and careful contrast stretching removes all atmospheric haze. Moreover, as noise increasingly dominates the higher order PCs, especially the last one (Fig. 5.23f), noise reduction is possible. Using median filters (Appendix B) on noisy PCs before the inverse

transform is one method. Because the highest PC rarely contains any meaningful information, simply omitting it from the inverse transform can achieve a dramatic reduction in the speckle and striping that result from sensor malfunctions, with little loss of spectral information.

5.5 Band ratioing

Just as the data in an image can be displayed as grey-tone images of single bands, or as three-band colour images, so can various arithmetic combinations of bands. The most useful of these is the ratio of one band to another. It is prepared simply by dividing the DN of each pixel in one band by that for the corresponding pixel in another band. Theoretically this should produce a range of new values for the pixel from zero to infinity, because some DN will have zero values. Because of the restricted range in each band of data and the strong correlation between bands, in practice ratios seldom fall outside the range 0.0–4.0. Displaying an image of the ratio requires rescaling these floating-point values to the 0–255 range of integers.

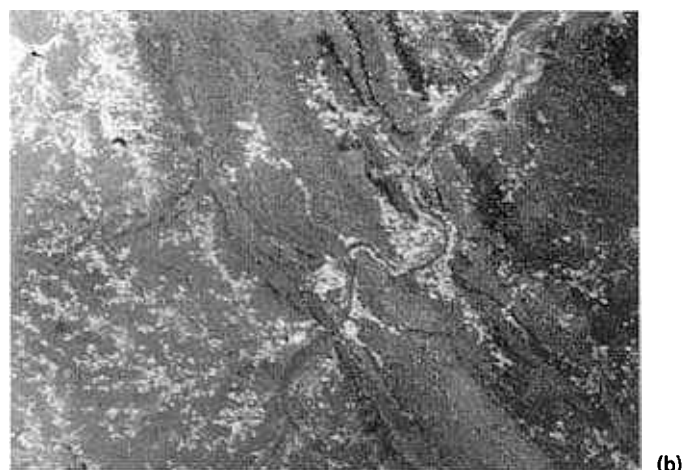
Random noise in different bands generally is uncorrelated, so ratios for pixels with spurious DN in either the numerator or denominator band produce extreme values. This enhances noise and most ratio images suffer from strong speckling and striping. They can be ‘cleaned up’ using the digital processes described in Appendix B. The contribution from atmospheric effects (Section 1.3.1) varies with wavelength, which means that ratioing also tends to accentuate its degrading effects. So, before ratios are calculated the bands involved should be corrected for atmospheric scattering (Section 5.2).

Individual bands of remotely sensed images show the effect of varying illumination caused by topography. The Sun illuminates a flat surface homogeneously. Where there is any appreciable relief, the lighting of slopes that face the Sun is stronger than for horizontal surfaces, and slopes facing away from the Sun receive less radiation. As a result, a surface with uniform reflectance properties will show varying DN across a scene. In a scene with many different surface types this can lead to added confusion in both detailed visual interpretation and in computer-assisted classification (Section 5.6.1) that use bands alone.

Theoretically any surface should receive the same proportions of energy in all wavebands irrespective of its orientation to the Sun. It also should reflect energy in the proportions controlled by its spectral reflectance properties. Therefore the ratio between two bands for pixels containing the same kind of surface should be the same, irrespective of the slope direction. For this to be true the surface must reflect radiation equally in all directions. Under natural conditions this behaviour is uncommon



Fig. 5.25 An image of enhanced Landsat MSS band 7 data for an area in South India (a) shows patches of strong reflectance that could be either bare soil or healthy vegetation. Displaying an image of the band-7 : band-5 ratio (b) exploits the spectral properties of vegetation to highlight vegetated areas and to



give some indication of the proportions of plant cover to bare soil. Pixels devoid of vegetation have the lowest value of the ratio and appear dark, whereas the brightest pixels have dense green vegetation cover.

and reflection reaches a maximum in a direction controlled by the structure of the surface and the angle of illumination. Another disruptive factor is the effect of illumination by diffuse radiation from the atmosphere; so-called 'skylight'. Radiation from this source, which is responsible for the illumination in deep shadow, has a variable contribution to the whole scene. Its spectral range is different from direct solar radiation, because it stems from wavelength-dependent scattering. Consequently, it changes the ratios on differently orientated slopes in a subtle fashion. Nevertheless, a ratio image does reduce the effects of slope and shadows to a marked degree (Plate 5.6a), and atmospheric correction before ratio generation reduces the effect considerably. Very much the same arguments hold for surfaces that vary in their albedo, but have similar spectral properties. They will appear very differently in images that use bands alone, but are expressed in the same way in a ratio image (Fig. 5.25).

Irritatingly, in these advantages lie serious disadvantages of ratioing. In its 'ironing-out' of topographic effects and the suppression of differences in albedo, ratioing hides important information. However, image processing is one endeavour in which it is possible to get a free lunch, or maybe even two. Using the HIS transform (Section 5.2.2) to convert the colour in a three-ratio image into saturation and hue, the 'lack' of spatially important information and the wealth of inevitable noise ends up in the intensity component. One ruse to restore spatially important information replaces the intensity from the HIS transformation of three ratios with a suitably stretched and edge-enhanced band. Plate 5.6b shows an example of this technique.

The most important property of a ratio image is that it

accentuates features in the spectral signature curve of a particular surface material. If the bands used cover peaks, absorption troughs or changes in slope of the spectral curve, then, combined in pairs as ratios they express well aspects of a material's spectral signature. The simplest example relates to vegetation. The characteristic features of plant spectra are the chlorophyll absorption feature in red and the strong reflection of infrared by living cells in the 0.7–1.2 μm VNIR range (Section 1.3.3). In images of red and VNIR radiance, pixels that contain a high proportion of living plants show as relatively dark and light, respectively. In a standard false colour composite image they will stand out from bare soil or rock as strongly red-coloured areas. Within these fairly obvious areas may be all sorts of variations. The proportion of a single species relative to unvegetated soil may change from place to place. Soil moisture or other environmental variations may affect the health and spectral properties of the species. The type of vegetation may vary, and there may be all manner of combinations of these important factors. Images expressing only data from various wavebands often conceal these differences, simply because the human eye has a limited range of perception. However, surface variations do change the structure of the spectral reflectance curve. In particular, the depth of the chlorophyll absorption feature, the peak height of VNIR reflectance and the gradient of the curve between red and infrared will change. Quite small changes in these parameters have a disproportionately large effect on the ratio of VNIR to red radiance (Fig. 5.25). The changes become easier to evaluate on the ratio image. Vegetation studies use ratio images of different kinds to estimate the proportion of green leaf cover and to help in discriminating between and classifying different plant species.

1 Ratios and mineral mapping

In a crude sense, ratios between bands describe the spectral 'colour' of an object, although colour perceived by humans corresponds only to the visible range. An important geological example does relate to perceptual colour. Characteristically, the Fe-O charge-transfer transition (Section 1.3.2) is a broad absorption feature at wavelengths less than 0.55 μm . It is responsible for strong red coloration of rocks rich in iron oxides and hydroxides. Sometimes this coloration is masked by mixing of iron minerals with large amounts of other minerals, which reflect strongly at all wavelengths, such as quartz. The albedo of such a mixture will be so high that it appears white in a natural- or false-colour image. However, the ratio of red to blue reflectance will enhance the small contribution of iron minerals, giving pixels on iron-bearing rocks a higher value than those composed of pure quartz. In Landsat TM data this measure of 'redness' corresponds to the ratio between bands 3 and 6.

Moreover, the position of the Fe-O feature is different for hematite (Fe_2O_3) and goethite ($\text{Fe}(\text{OH})_3$), hematite being more absorbent in green than is goethite, which results in the former being cherry red and the latter an orange brown. Potentially the red/green ratio can discriminate between the two minerals; goethite is the most common weathering product of pyrite (FeS_2) in mineral deposits. Iron also produces an absorption band between 0.85 and 0.92 μm , owing to an electronic transition (Section 1.3.2). This feature falls within the range of Landsat TM band 4, whereas a reflectance high for all minerals is found in band 5. The ratio of band 5 and 4 therefore shows higher values for oxidized iron-rich rocks than for other types. The Al-OH and Fe-OH rotational transitions associated with clays and other hydroxylated minerals result in absorptions within Landsat TM band 7. Dividing this band by band 5 results in clay-rich rocks showing up as dark areas.

Using Landsat TM data these four spectral features define the practical limit to surface discrimination using ratio techniques, even though there are 30 possible ratio combinations of the six reflected bands. Moreover, the wide band widths mean that only gross differences are detectable. Many other features can be investigated using narrower bands within the same reflected range, as from SPOT-1, ASTER and hyperspectral devices (Section 5.6.2). To introduce the potential of the technique, Fig. 5.26 shows the ratios of Landsat TM 5/4, 3/1 and 7/5, together with a false-colour band to show spatial features of the area of study. Figure 5.6 shows them in this order as an RGB image.

2 Ratios and geobotany

Plants depend on the underlying soils and rocks for the supply of water and nutrients. In their natural state

plant communities are sometimes excellent indicators of geology in humid terrains. In some cases, specific plants provide favourable chemical conditions for growth of restricted plant communities, such as calcium- and magnesium-loving species, and even species that will not grow over rocks rich in transition metals. Sometimes the substrate threatens the health of plants. This may be due to the result of anomalously high concentrations of certain elements, as might be the case in the vicinity of certain mineral deposits. More commonly, however, such stressed vegetation is merely a response to either watering or a shortage of water, in which case it generally represents a seasonal phenomenon. Different plants have different tolerances to environmental stress, so another ecologically related feature of vegetation is the interaction of species that are favoured or stressed by particular geological conditions. Geobotanical indicators can be examined most successfully using ratio techniques based on careful reference to field reflectance spectra.

Figure 5.27 shows how the reflectance spectrum of a plant changes with the moisture content of its leaves. The progressive deepening of the absorption features at 1.45 and 1.9 μm as leaf moisture increases is very clear. Although atmospheric effects mask this portion of the spectrum, the related changes in the regions around 0.8 μm relative to the shorter infrared plateau, and in the 1.6 μm region, show that ratios can be devised to highlight variable moisture stress. With Landsat TM data the 4/5 and 5/7 ratios should increase as leaf-moisture content increases. Geochemical stress, either through shortages of essential nutrients or the presence of toxic compounds, sometimes results in the near absence of vegetation over the offending rocks and soils (Plate 5.7). A common result is the stunting and discoloration of plants growing on them. Disruption of the pigmentation in plants produces shifts in the position of the chlorophyll absorption features associated with the red and blue ends of the spectrum. It also may affect the shape of the spectrum near to these features. These signs of geochemical stress are often most discernible at particular stages of plant growth, commonly just before the onset of leaf senescence (Fig. 1.17). At present the wavebands most readily available from orbital systems are too broad to detect these changes. However, the shifts and distortions are definable by ratios between very narrow bands of hyperspectral data.

Pattern recognition

Research in military and various scientific fields since the late 1970s has developed several mathematical means for automatically extracting information from digital images. The techniques fall under the general heading of artificial intelligence. They were devised originally to

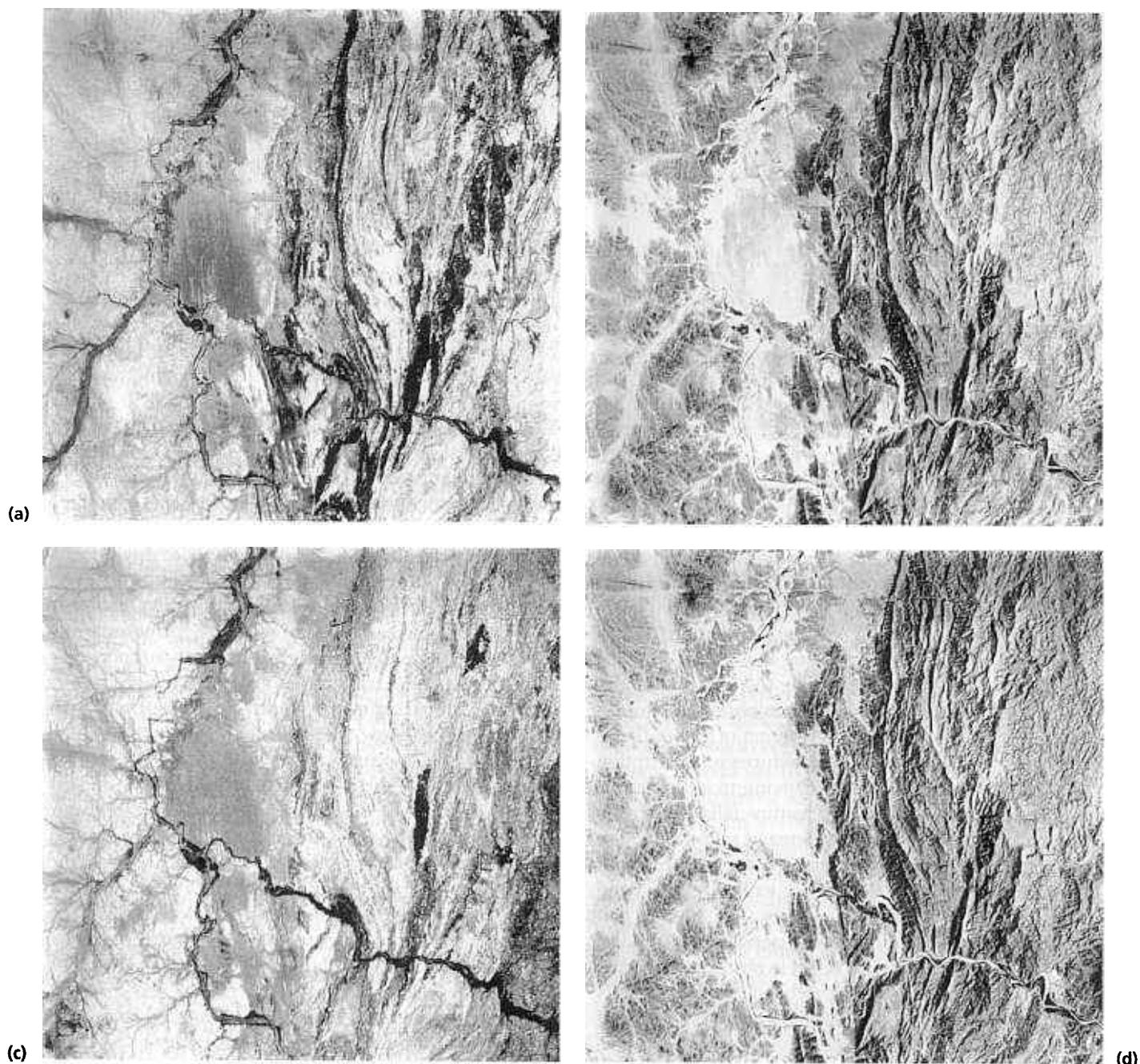


Fig. 5.26 Images (a) to (c) show the TM ratios of bands 5/4, 3/1 and 7/5 compared with a band 3 image (d) for the area on the Sudan-Eritrea border used in Fig. 5.4 and Plate 5.1. See also Plate 5.6.

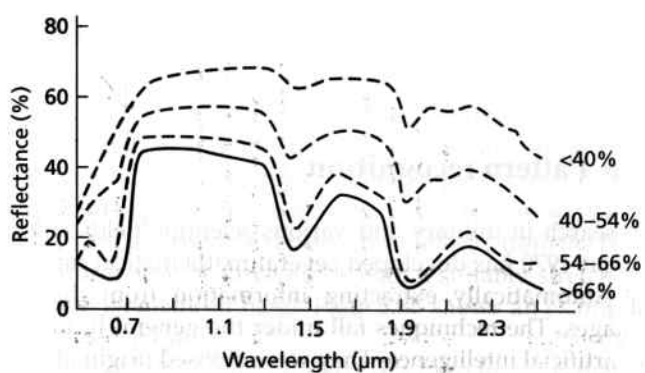


Fig. 5.27 (left) The leaves of Indian corn have reflectance spectra that show considerable changes with increasing moisture content. Similar curves have been obtained from various plant species.

aid in the recognition of specific objects, such as tanks or engineering components, to speed up intelligence gathering and develop robotic vision. Such pattern recognition is focused on three main aspects. First is the detection of different *kinds* of surface materials, such as camouflage or colour-coded objects, which relies on the tonal or spectral characteristics of images. Most cheaply available digital data have limited numbers of spectral bands, which generally are so broad that they span several spectral features. Nonetheless, different surface materials do show sufficient variability in their broad-band reflectance or emittance for them to be distinguished visually on suitably enhanced and transformed images, as described in Sections 5.2, 5.4 and 5.5. The differences can be systematized and exploited digitally in semiquantitative classification. If hyperspectral data are available, the coverage of visible to SWIR wavelengths by many narrow bands produces near facsimiles of spectra for each pixel in the scene. This opens the possibility of spectral pattern recognition and mapping variations in surface cover quantitatively (Section 5.6.2). Of equal importance is the discrimination of distinctive shapes, such as warships or tanks, from natural spatial patterns. Third is the recognition of changes in an area with time, as in the case of the progressive build-up of military forces or the redeployment of missile silos. Although most of the software involved stemmed from the military, medical and engineering communities, much of it is very useful in environmental science applications.

As geological features are so variable and because they are often masked by soils and vegetation, visual interpretation, even by an experienced person, is often ambiguous. This is made even more of a problem by the sheer number of theories that a geologist can use in assessing the context of features in an image. Artificial intelligence and pattern recognition open up the possibility for

geologists to 'train' computers in making some geological decisions in a less biased way.

5.6.1 Multispectral classification

Automatically discriminating surface materials on digital images using computer software, or classification, is based on statistically separable differences between their radiances in the available bands (Fig. 5.28). Such differences stem from the materials' spectral properties (Fig. 5.29), which hyperspectral images record almost faithfully, and which are the most reliable means of quantitative discrimination (Section 5.6.2). Hyperspectral data are, however, a far cry from the most cheaply and easily available multispectral images that use broad bands. In them the subtle features of a material's spectrum are degraded to average DN for widely spaced portions of the full spectral range, even though the wavebands aim at highlighting the main spectral features of specific materials.

Inspection of Figs 5.28 and 5.29 suggests that a computer could perhaps achieve a good discrimination between some of the different surface types using data from one band alone. For instance, water has very low DN in band 10 compared with vegetation and all rocks. Although there are also differences among the rock types, and between them and vegetation in all the bands, there is no guarantee that there is not considerable overlap. The bars and spectral curves relate only to specific localities within a scene. Over the whole scene the reflectance properties of different surface categories vary within limits determined by natural variations in the materials themselves and mixtures between materials. Figure 5.30 shows that the overall histogram for any one band is built up from individual ones that represent different materials. Many of these 'subhistograms' overlap.

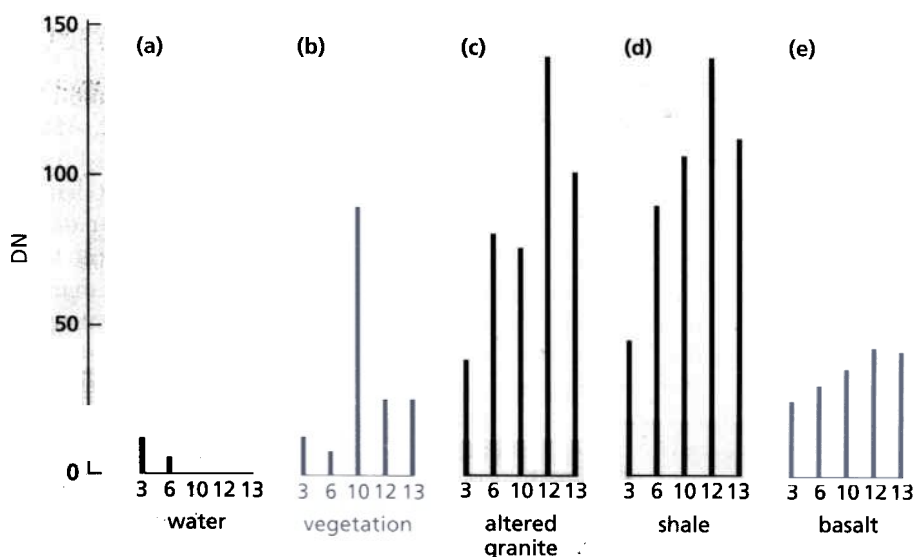


Fig. 5.28 Spectral reflectance curves can be represented crudely by bar charts corresponding to digital numbers (DN) in the restricted number of wavebands covered by most imaging systems. In this case (a) represents water, (b) vegetation, (c) hydrothermally altered granite, (d) shale and (e) basalt. The waveband numbers shown are those in a multispectral line-scanner.

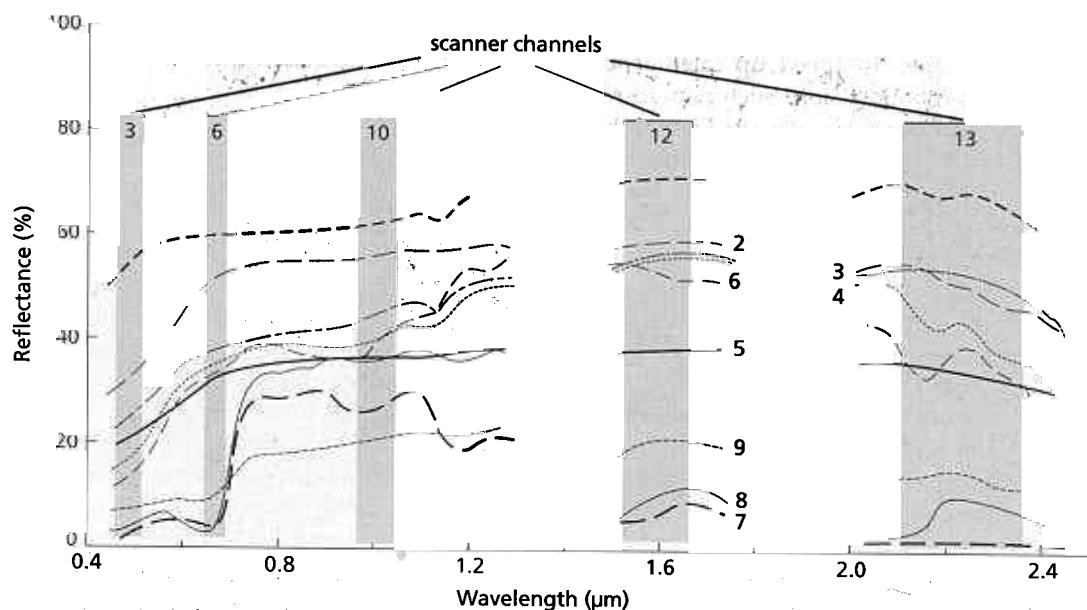


Fig. 5.29 Superimposed on the reflectance spectra of a number of surface materials from Utah, USA are the wavebands used in Fig. 5.28. From Podwysocki *et al.* (1983).

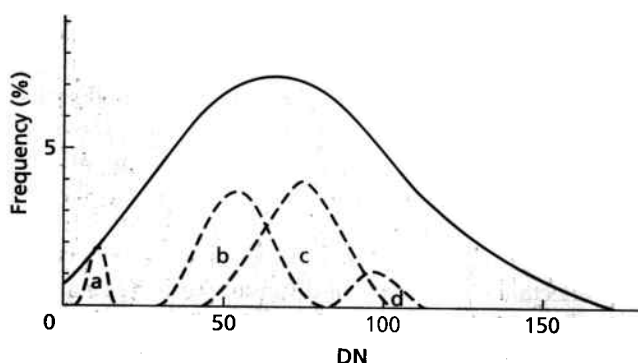


Fig. 5.30 The histogram of one band for a whole scene conceals within it smaller histograms for particular classes of surface. In this case (a) to (d) are hypothetical classes.

Some are narrow and others are broad. Each may be described by the mean DN for the category that it represents and the variance, which expresses the spread of DN within a category, and is considered shortly. The differences can be enhanced using ratio techniques (Section 5.5), where the high VNIR reflectance relative to that in red for vegetation distinguishes it uniquely from other natural materials. Things are never as simple as they might seem, however. Pixels themselves form the limit to discrimination, and frequently are so large that they represent mixtures of exposed materials in small patches that comprise pixels on the ground (Fig. 3.10). Even with near-continuous spectra, a surface class is likely to be 'impure'. That is bad enough for categories of vegetation, but worse for geological materials because of two conspiring factors. The first of these is that remotely sensed information, from whatever part of the electromagnetic spectrum, emanates from the surface only,

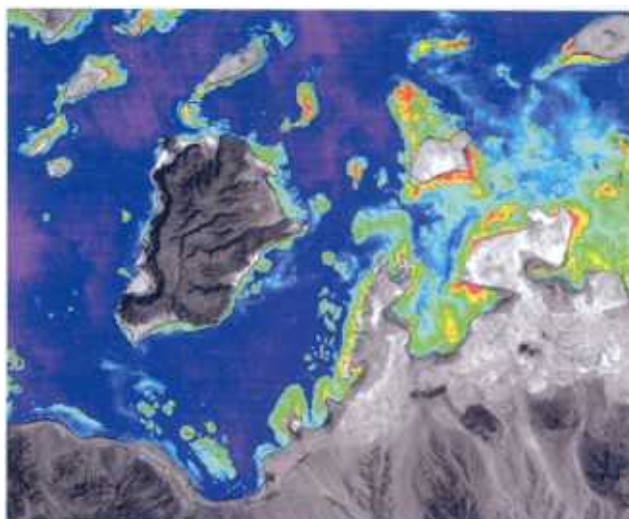
which almost inevitably is weathered to some extent. Rock-forming minerals are replaced by their altered products, at least to some degree. Imagine a granite formed from feldspar, quartz and muscovite. Its weathered surface will be quartz with unweathered feldspar and mica, plus varying proportions of clays formed by feldspar weathering, let alone the minute but highly coloured products from conversion of iron compounds in the rock. Rocks rich in ferromagnesian amphiboles, pyroxenes and olivines will have surfaces full of clays, serpentines, chlorites and oxidized iron minerals. Although rocks are divisible into the familiar classes using a hand lens or microscope, even if they are weathered a little, petrography is impossible from more than a few metres away!

The second difficulty is this. Except in the most arid terrains with little superficial cover, rock outcrops are really quite rare, though often very prominent. They are mantled to various degrees by soil, transported sediments and vegetation (Fig. 5.31). Although vegetation and soils may be related to rock type, even uniquely in some cases, the geologist requires an outline of where different rocks are present beneath the veneer—that is what a geological map sets out to show. Even in areas of complete exposure, classification faces problems. In deserts, thin surface veneers of manganese and iron oxides known as desert varnish coat rock surfaces, which, irrespective of their composition, appear dark in visible and near-infrared images. A similar problem is encountered with highly reflective encrustations of salts formed by evaporation. In some areas of diverse vegetation and soil cover the only possibility in classification is to select subclasses within each lithological category, such as bare rock, derived soil and vegetated surfaces with different degrees of plant



(a)

Plate 1.1 The clear water of the Red Sea off Mersa Fatma in Eritrea allows green light to penetrate to depths of 20–30 m. In image (a), a Landsat TM false-colour image, the blue colours of the nearshore waters are to the result of reflection from coral reefs and sand in shallow water. Image (b) shows colour bands



assigned to narrow ranges of blue reflectance to produce a representation of the variation in bathymetry. Grey-tones are above sea level, red is the shallowest water and the 'rainbow' progression towards dark blue then magenta is from shallow to deep water.



Plate 2.1 The huge range of colours has been produced by combining the three additive primary colours in the proportions shown as red, green and blue. Courtesy of Adobe Systems Inc.



(a)

Plate 3.1 In this pair of vertical aerial photographs of a rural scene in Devon, UK (a) is natural colour and (b) false-colour infrared. Although the natural colour image is more pleasant to look at, because it is dominated by wavelengths near the peak of human visual sensitivity, it contains less information than the false-colour infrared image. This is particularly striking in the wooded area, where not only are the



(b)

conifers sharply distinguished from the sparse, bright red broad-leaved species in the dark swath of trees, but many stands of different broad-leaved species, different stages of growth and different canopy types can be separated too. On the natural colour image these features are by no means obvious. Copyright Aerofilms Ltd.

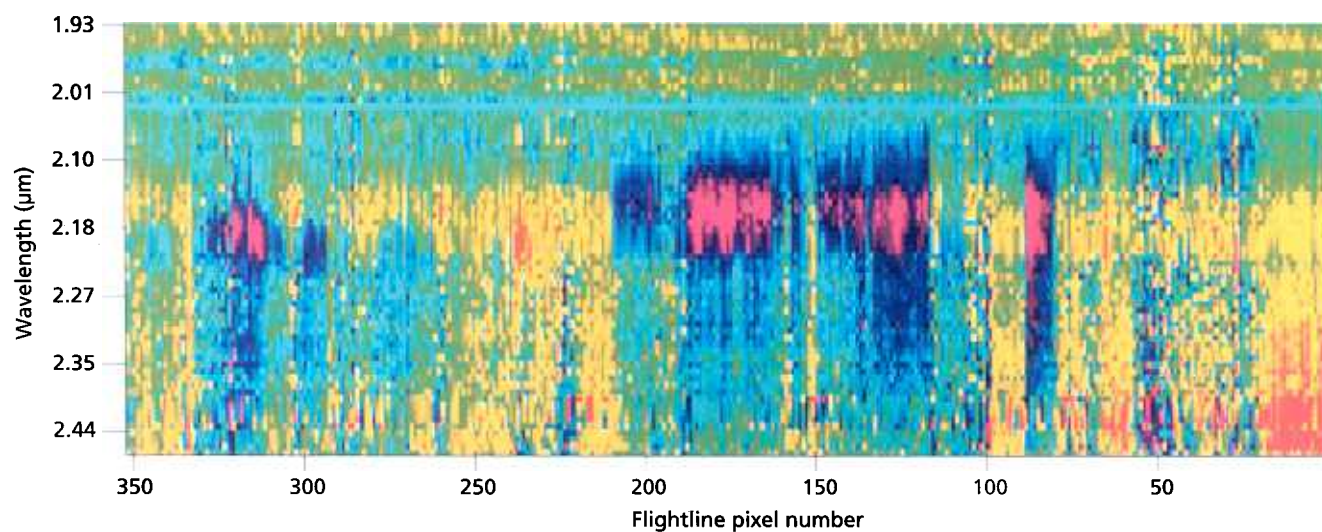
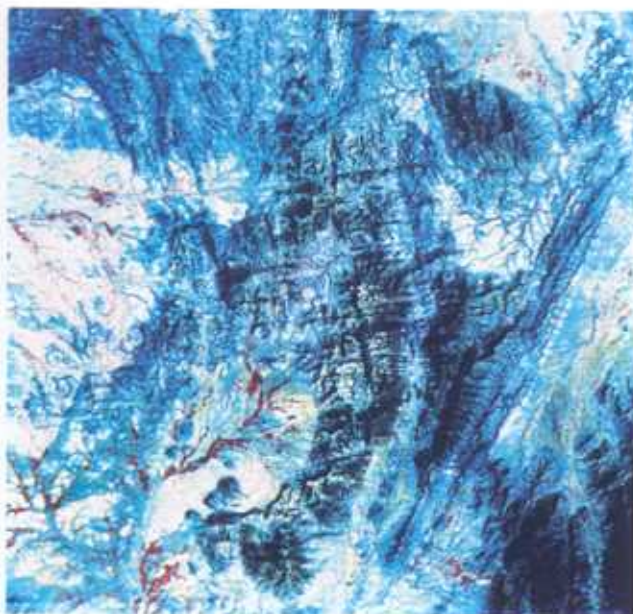


Plate 3.2 This figure consists of many spectra from an airborne spectroradiometry campaign. The spectra are stacked for individual ground resolution cells along a flight line from right to left. The vertical axis spans wavelengths from 1.93 to 2.44 μm in 20 nm bands. For each narrow spectral band the spectrum for a single ground resolution cell is colour coded from red (high value) through a rainbow sequence to violet (low value). The colour patterns indicate the distribution

along the flight line of peaks and troughs in the surface reflectance spectrum. These in turn may indicate precisely the abundance of minerals that have distinctive absorption bands. The magenta patches between 2.1 and 2.35 μm suggest abundant clay minerals (Fig. 1.9) at the surface from samples 70–80, 120–210 and 300–330 along the flight path. Courtesy of Stuart Marsh, Sun Oil, Houston.



(a)



Plate 3.3 (a) Landsat-2 MSS false-colour image for part of Eritrea (see Fig. 3.28). Image (b) shows enhanced Landsat-5 TM data for the same area, where red is band 5 data, green is band 3 data and blue is band 1 data. The 28.5 m resolution of the TM data produces a sharper image than the MSS data, helping

delineate finer drainage and topographic textures. The extension of the range in the near-infrared by TM allows variations in surface composition to be expressed in a wider range of colours than is possible with MSS data.



Plate 3.4 Mosaic of cloud-free AVHRR images showing Western Europe and part of North Africa in a simulated natural-colour rendition. Courtesy: British National Space Centre.



Figure 4.1 The Colorado River of Arizona is a classic example of antecedent drainage. In this false-colour Landsat TM image at a scale of about 1 : 3 million it has carved the spectacular Grand Canyon through the rising Colorado Plateau because its rate of downward erosion exceeded the rate of crustal uplift.



Figure 4.3 This enhanced false-colour composite of Landsat data shows an area in the central part of the Tibetan Plateau. Although it is of marginal use in this form for discriminating lithologies, the many linear features that it displays are the key to understanding the region's active tectonics. Several of them clearly displace Recent superficial deposits, as at centre right and centre left. Scale 1 : 1 million. Source: courtesy of Dave Rothery, Open University.

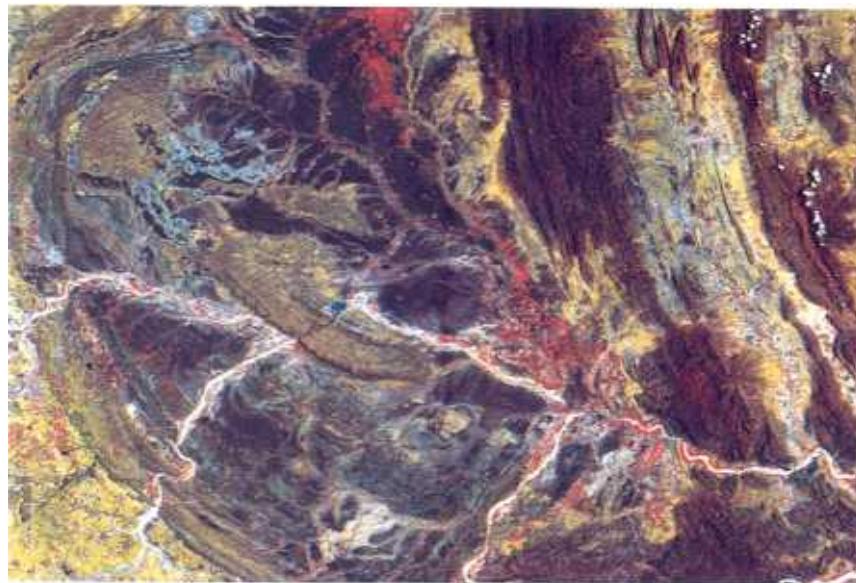


Figure 4.2 The curved dark bands on this enhanced Landsat false-colour composite of part of Andhra Pradesh in central India are outcrops of sediments that filled the huge Godavari Basin between 1.7 and 1.3 Ga ago. The yellow area in the south-west comprises 2.6 Ga Archaean granitic gneisses. Dykes cutting the Archaean show up very clearly as

secting dark lines, but stop at the boundary with the late Proterozoic amphibian rocks. This boundary is clearly an unconformity. Radiometric dates on the dykes show that they are about the same age as the oldest of the basinal sediments. Scale: 1 : 1 million. Source: author, written to film by Michael Smallwood, Earthwatch Inc.



Plate 4.4 The western part of the Qaidam Basin in Qinhai Province, China, shown on this Landsat MSS false-colour composite, is floored by folded lacustrine and other nonmarine sediments of Tertiary to Quaternary age. They occupy the mainly grey texturally flat area. The basin is bounded to the NW by the sinistral Altun strike-slip fault, which forms a prominent linear valley. The rugged area in the north is an uplifted massif of older, Mesozoic rocks. Scale: 1 : 1 million. Source: author.



Plate 4.5 Superimposition of intense shear strains and upright folds on an earlier set of recumbent folds has produced several fold interference structures in a Precambrian metamorphic complex in north-east Sudan. They show up well on this enhanced Landsat TM false-colour composite because the different lithologies are strongly contrasted in colour. The image is 30 km across. Source: author.

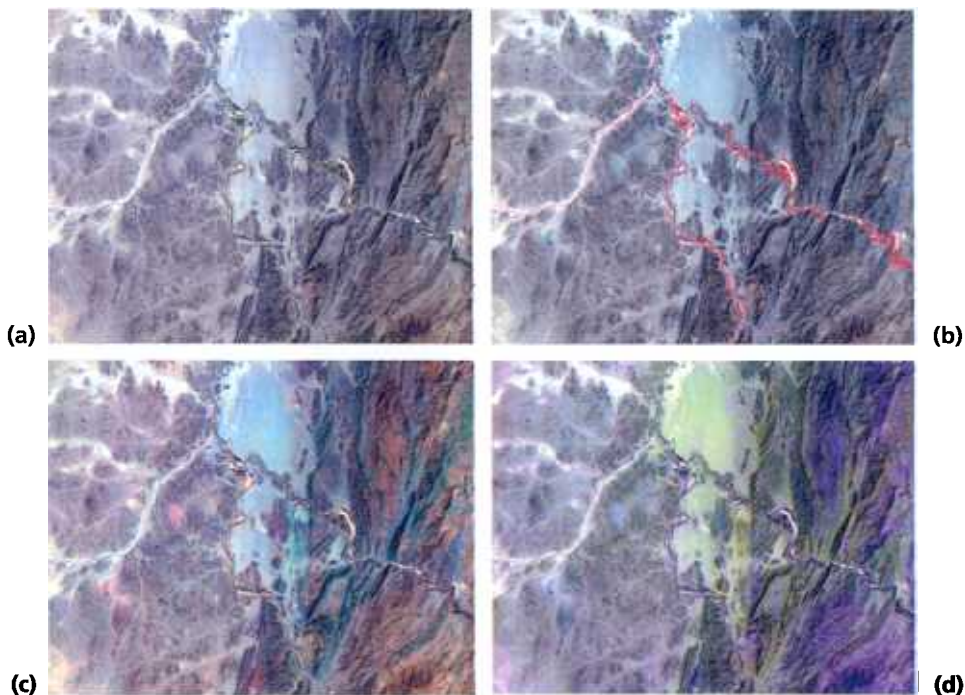


Plate 5.1 These linearly stretched false-colour Landsat TM images of part of the Sudan-Eritrea border are of the area shown in Figs 5.4–5.9. They reveal more details than the band 3 images in both light and dark terrains, but each band combination contains different information. The natural colour image (bands 3, 2 and 1 as RGB) (a) is least informative. The standard false-colour composite (bands 4, 3 and 2 as RGB) (b) shows vegetation in shades of red, but little more than (a). Combining bands 5, 3 and 1 in RGB (c) reveals better detail of the distribution of different rock types, while suppressing vegetation as dark tones. The importance of a correct

assignment of bands to colour components is shown by (d), which has bands 3, 1 and 5 in RGB. Exactly the same information as in (c) is shown, but the choice of colours prevents the eye from appreciating the full range of variability. The geology is of a major boundary between a high-grade metamorphic terrain with granite intrusions and basaltic dykes to the left, and a very complex, younger terrain of peridotites, metabasalts and metasediments to the right. The boundary is marked by deep alluvium along a major river controlled by a huge shear zone. Each image is 30 km across.



Plate 5.2 Cross-section through the conical representation of intensity-hue-saturation (IHS) colour space, or colour wheel, showing how pure spectral hues on the perimeter grade through pastel shades to grey at the axis of the cone. In this case red is 0 at the top.

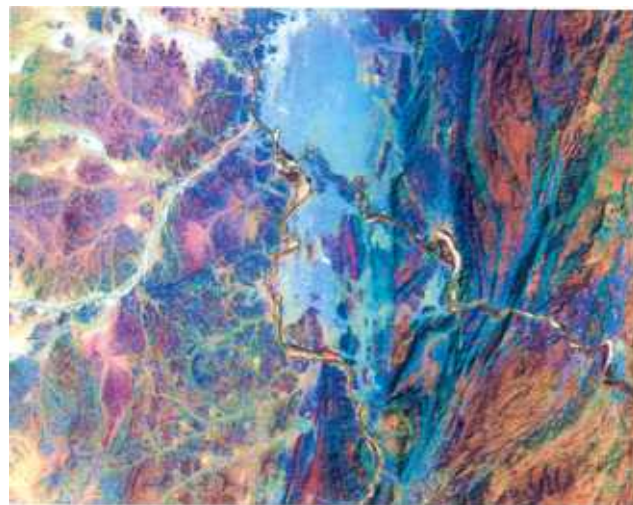


Plate 5.3 The Landsat TM band 5, 3 and 1 data in Fig. 5.10(c) have been transformed to intensity-hue-saturation (IHS) co-ordinates, with the intensity and saturation files stretched and the hue unchanged. The results have been transformed back to RGB space to produce a vivid range of colours. Comparison of the two images reveals that the basic colour differences in the image have not been changed, so that colour can still be related easily to spectral properties.



Plate 5.4 This image combines the first-, second- and third-order principal components from the area of Plate 5.1 and Fig. 5.10 in red, green and blue. The result is a vivid colour representation of surface variations in which much geological detail is visible, particularly different kinds of rock. The colours are not easily related to the rocks' spectral properties, however, and the compression of most of the variance into the first principal component (PC) means that the other PCs are noisy and noise interferes with interpretation in the false-colour image.

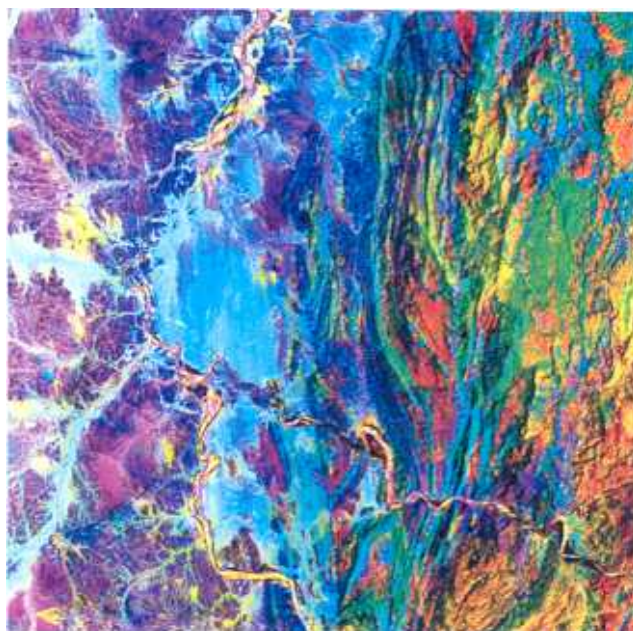
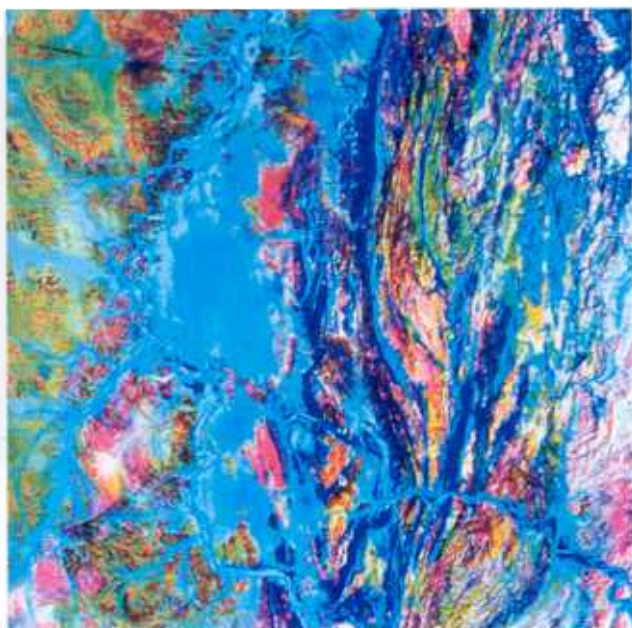


Plate 5.5 Decorrelation- or D-stretched image of Landsat TM bands 5, 3 and 1 in RGB for the same area as Plate 5.1 (c). The first principal component (PC) was stretched and edge-enhanced, and higher order PCs filtered to remove noise before rotation back to original band space. It shows little difference compared with Plate 5.3, but the D-stretch technique offers more opportunities for cosmetic improvement than the intensity-hue-saturation (IHS) transform.



(a)

Plate 5.6 (a) False-colour ratio image of the area of Fig. 5.26, using TM 5/4 as red, 3/1 as green and 5/7 as blue. Although producing an excellent discrimination of rock types, the image lacks some topographic detail and is noisy. By using the intensity–hue–saturation (IHS) transform and substituting a



(b)

contrast stretched TM band for the ratio intensity (b) both defects can be avoided without changing the colour rendition of different rocks, which is dependant on spectral features related to iron minerals, albedo and hydroxyl-bearing minerals. Both should be compared with Plates 5.3 and 5.4.

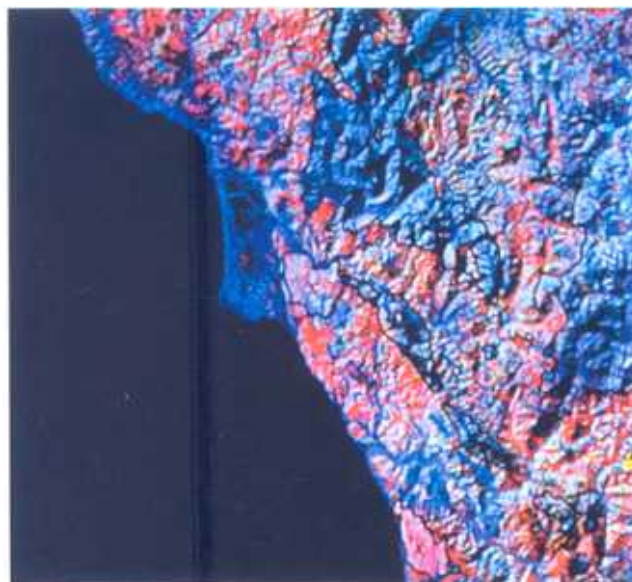


Plate 5.7 Vegetation in the humid climate of northern California is luxuriant over most rock types. However, ultramafic rocks produce soils that inhibit growth. On this colour composite of Landsat MSS band ratios—4/5, 4/6 and 6/7 as red, green and blue—the bluish areas have sparse vegetation compared with the red areas. They correlate very well with known outcrops of ophiolites—black outlines. Courtesy of Gary Raines, US Geological Survey.

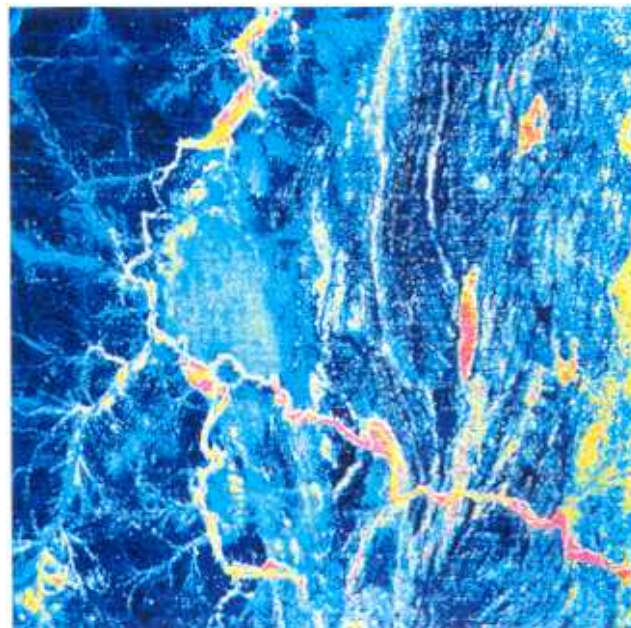


Plate 5.8 This image is a density-sliced version of the ratio of Landsat TM bands 5/7 (see Fig. 5.26). This ratio should be correlated with the proportion of hydroxyl-bearing minerals in surface rocks and soils. The colours are assigned to ranges of ratios, the 'cool' colours representing low hydroxyl-mineral contents, the 'warm' colours higher contents.

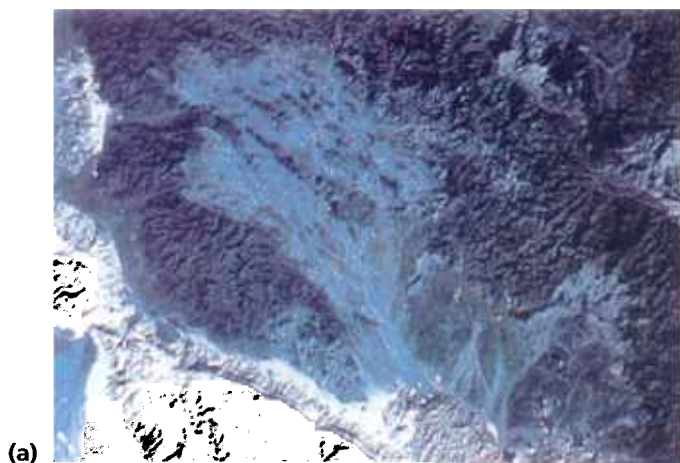
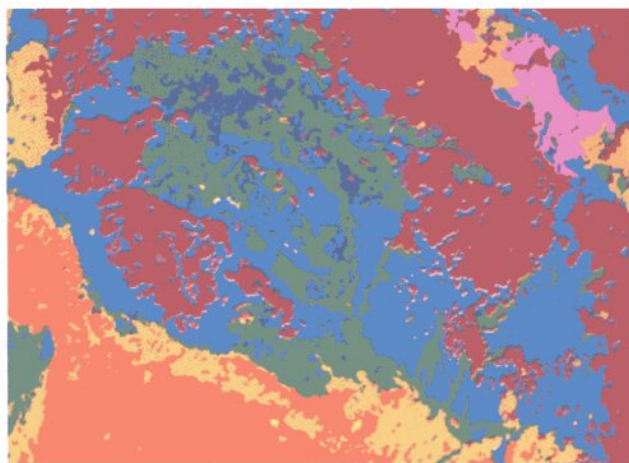


Plate 5.9 The Cretaceous ophiolite complex of the Sultanate of Oman is difficult to interpret from a false-colour Landsat MSS image (a). Most of the ophiolite is very dark. However, careful manipulation of the data and use of supervised maximum-likelihood classification techniques allows a crude discrimination between the different major lithologies to be made. In (b) the colours light blue, green and dark blue are



assigned to hartzburgite, layered gabbro and sheeted basaltic dykes, respectively, which comprise the ophiolite, the brown indicates zones of pervasive serpentinization in the hartzburgite, and the pink, yellow and orange relate to volcanic rocks and limestones, which are overlain tectonically by the ophiolite. Courtesy of Dave Rothery, Open University, UK.

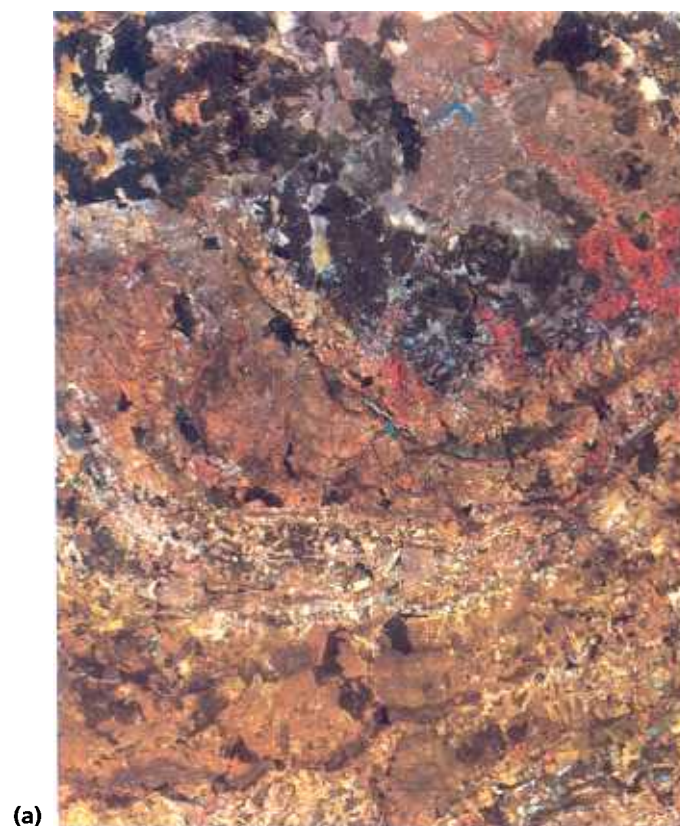
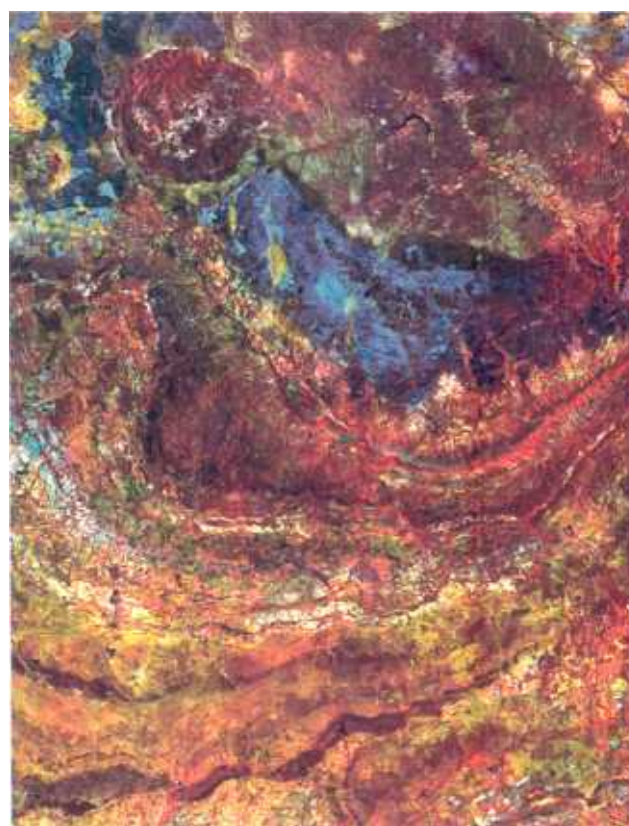


Plate 5.10 Southern Africa has a strongly seasonal climate, which has a marked effect on the appearance of remotely sensed images. Image (a) is an enhanced false-colour composite of Landsat MSS data for an area in the Transvaal, South Africa during the dry season. Although some geological features show up, they are muted by the effect of dead vegetation and dry soil. The same area imaged just after the



wet season (b) shows a spectacular improvement. This is the result of strong geological control over vegetation and variations in soil moisture. Important features are the Bushveldt layered basic-ultrabasic intrusion in the north-east, the circular Pilansberg intrusion at top centre and previously unsuspected stratigraphic variations in the Precambrian Transvaal dolomites (lower half). Courtesy of ERIM, USA.

Plate 6.1 This colour image of HCMM data is of the same area as that shown in Fig. 6.11. The red, green and blue components are controlled by negative images of Day-IR, Night-IR and Day-VIS, so a warm area during the day imparts a dark red to the image, a cold area at night contributes bright green, a surface with high albedo is represented by a dark blue addition, and so on. The resulting colours combine information on albedo and thermal inertia in a striking way. As well as a much better discrimination between many different major rock units compared with Fig. 6.11 (see Units 15, 2, 3, TK, Km) there are several important points to note. Sandstone Unit 10 shows up as blue. It has low albedo, and is warm during the day and cold at night, implying a low thermal inertia. This is related to its red colour and high silt content. Other sandstones, such as Units 12, 13 and 14, appear dark orange (high albedo, low daytime temperatures, warm at night, and hence high thermal inertia). They are pale, coarse-grained, massive sandstones. Courtesy of Anne B. Kahle, Jet Propulsion Laboratory, Pasadena.

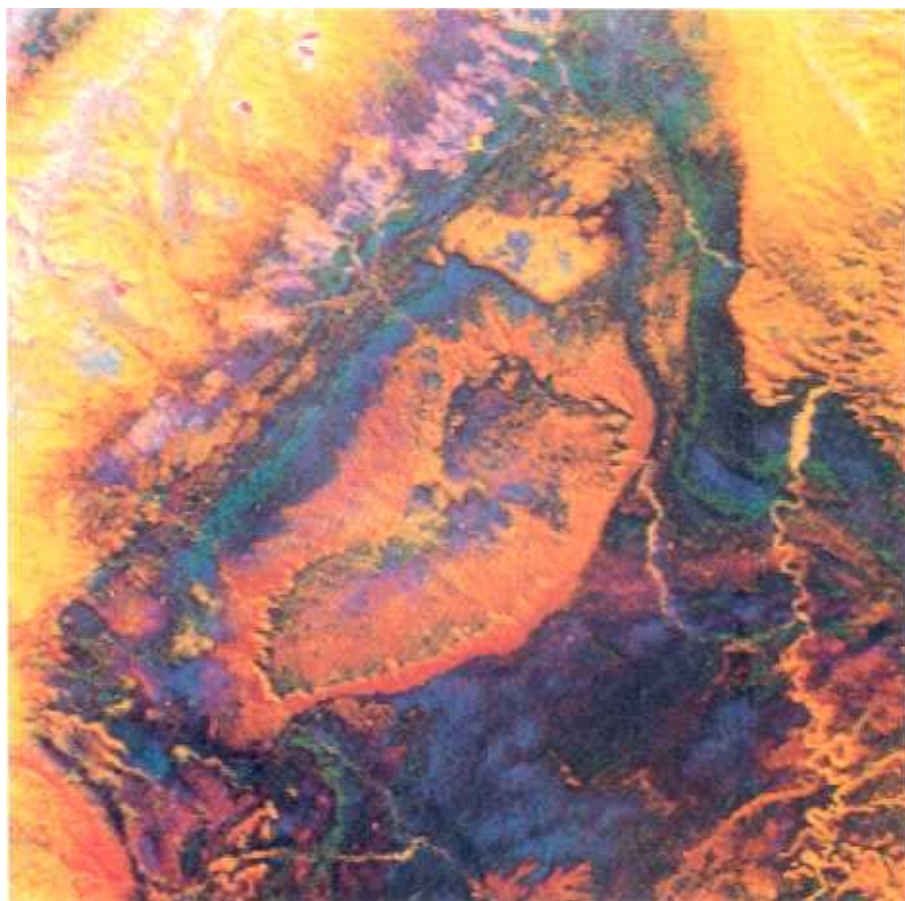
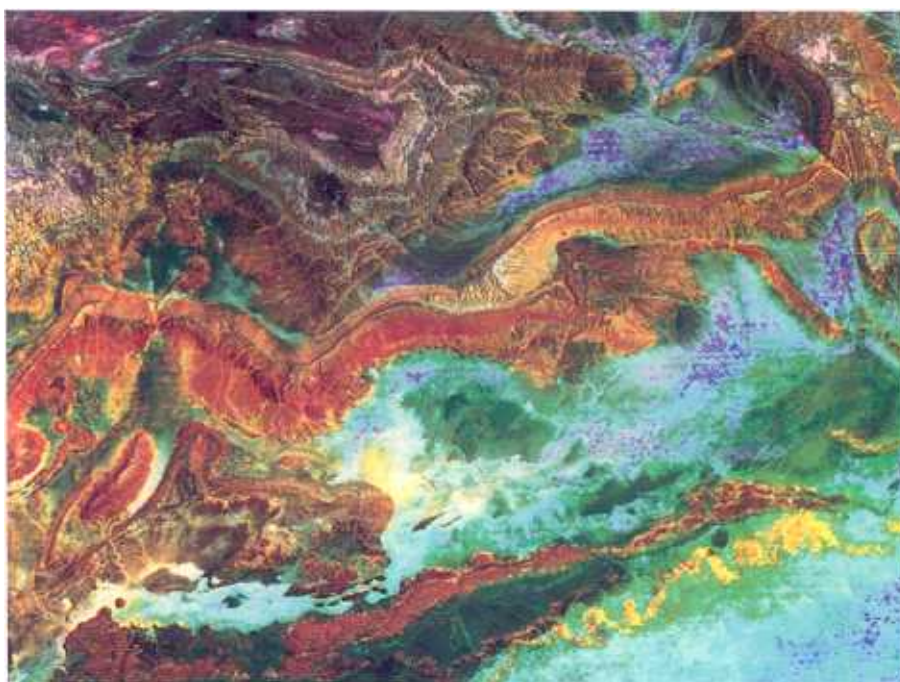


Plate 6.2 This image results from displaying Landsat MSS band 7 data (representing albedo) with HCMM Day-IR and Night-IR as intensity, saturation and hue, respectively. The area is of the High Atlas Mountains in Morocco, which are formed by a fold belt of Permo-Triassic sandstones and Jurassic to Cretaceous carbonates resting on a Precambrian and Palaeozoic igneous and metamorphic complex. The valleys are partially filled with Tertiary to Recent sands and alluvium. The various colours can be interpreted in terms of thermal properties by reference to Table 6.2. Courtesy of Rupert Haydn, University of Munich.



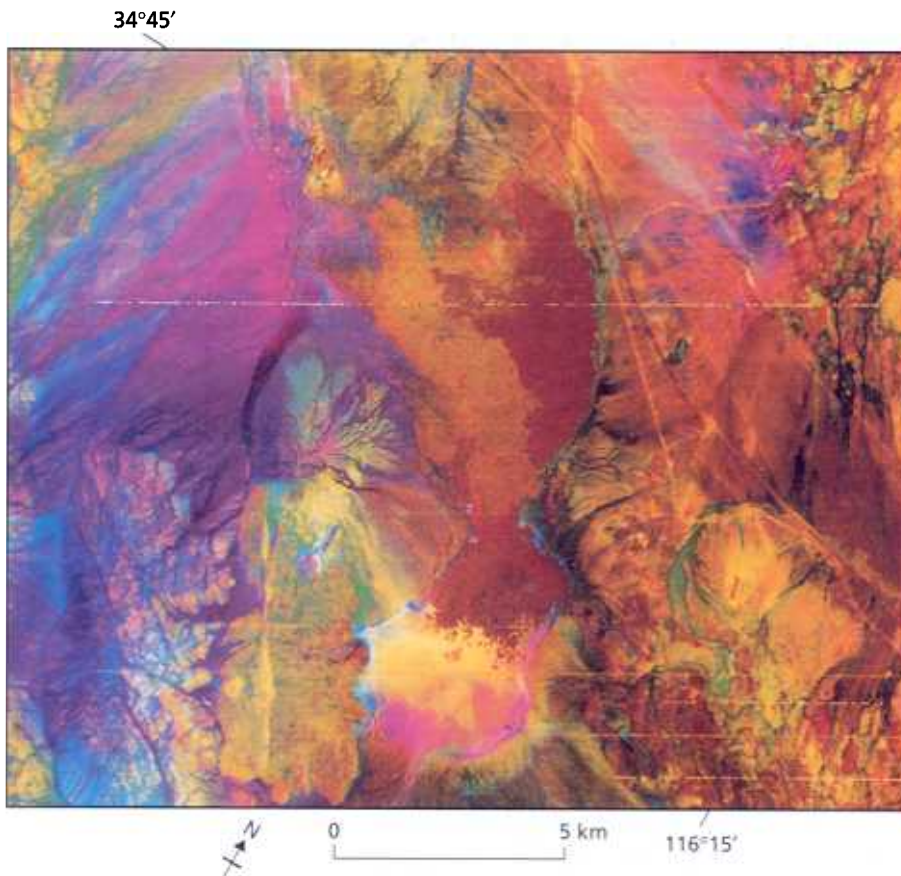
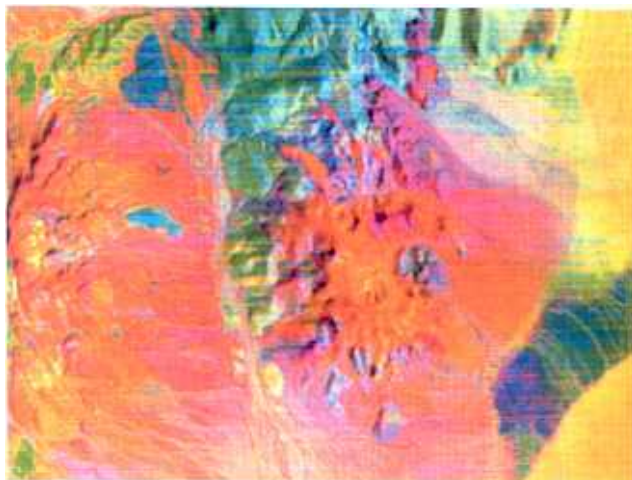


Plate 6.3 Combining data from visible, near- and thermal-infrared parts of the spectrum through principal component analysis and RGB display results in spectacular colour images. In this case, components loaded with visible and near-infrared data control red and green, whereas blue represents a component containing most of the day and night thermal information. It covers the same area as Figs 6.7 and 6.12, Pisgah Crater and associated lava flows in the Mojave Desert of south-eastern California, USA. It is impossible to relate image colours to a multitude of contributing surface properties, and the image must be interpreted empirically. The central dark-red and orange units are aa and pahoehoe flows from Pisgah Crater, which is brighter orange—it is a cinder cone. The yellow and magenta area at bottom centre represent a playa lake. Different igneous units are clearly discriminated in the lower left and lower right quadrants, and many subtle differences in the alluvium at the top left also show up well (see the map in Fig. 6.12). Courtesy of Anne B. Kahle, Jet Propulsion Laboratory, Pasadena.



(a)

Plate 6.4 Image (a) is a false-colour image of TIMS bands 5, 3 and 1, displayed as red, green and blue after contrast stretching of principal components and rerotation into the original data space (Section 5.4). Image (b) is a colour ratio composite of 1.6–2.2 μm (red), 1.6–0.48 μm (green) and 0.6–1.1 μm (blue) reflectances from an airborne multispectral scanner. Both are of the Cuprite mining district in Nevada, USA. The TIMS image is much better than the ratio image in distinguishing tuffs, basalts, carbonates and siltstones, plus examples of outcrops enriched in clay minerals. The ratio image is better for separating alluvial deposits. The area contains rocks that have suffered two different kinds of alteration during hydrothermal activity and mineralization,



(b)

in a roughly circular zone just left of the centre, which both images highlight in mainly red colours. In the ratio image clay-rich parts of the altered zone have a high 1.6/2.2 ratio and appear bright red. Those with a high iron content and a high 1.6/0.48 ratio appear green. Silicified rocks have lost their clay content and are dark red, brown and bluish. Yellow patches reflect zones rich in both iron and clay minerals. The TIMS image expresses these variations rather differently, and the colours are affected most by silica content—either quartz or opal. Silicified rocks are bright orange, and clay-rich opalized rocks are magenta. Both types of image complement each other in the search for this type of mineral deposit. Courtesy of Anne B. Kahle, Jet Propulsion Laboratory, Pasadena.

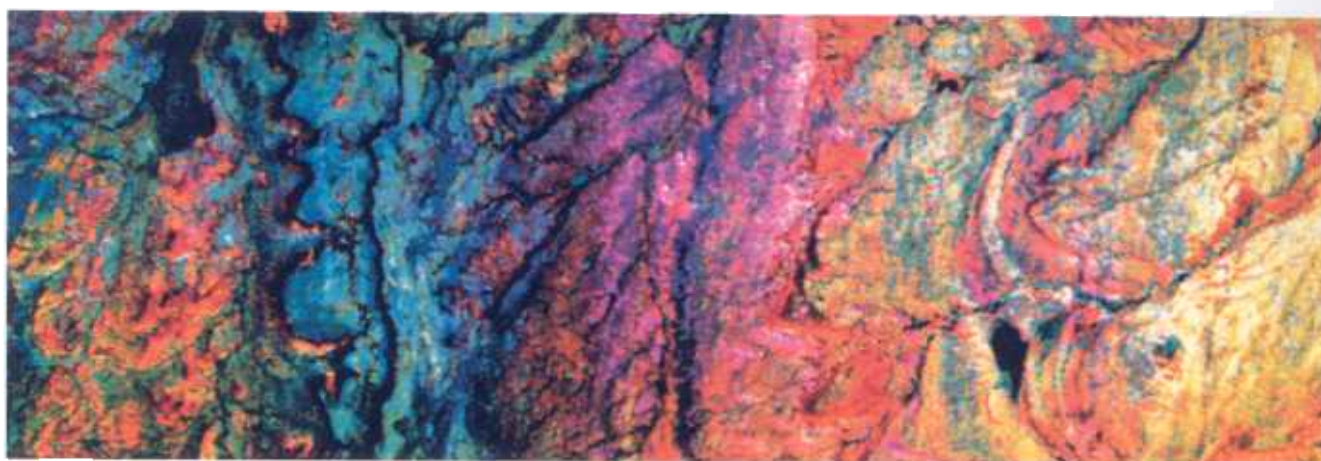


Plate 6.5 This TM image was produced in the same way as that in Plate 6.4. It is of part of the Mesozoic Wind River sedimentary basin in central Wyoming, USA. Because of their mid-infrared spectral features (Fig. 6.15), silica-rich sediments are portrayed in red to red-orange, and carbonates in green and blue-green. Compared with the Thematic Mapper image of the same area (Plate 7.1), many significant variations in stratigraphy, which are invisible in the reflected part of the

spectrum, are quite evident in this image. The blue-green carbonates contain sporadic patches and discontinuous layers of red sandstones. The brightly coloured, dominantly arenaceous rocks on the left owe their variation in colour (pinks, greens and yellows) to variable proportions of carbonate and silica in their cement. Courtesy of Harold Lang, Jet Propulsion Laboratory, Pasadena.

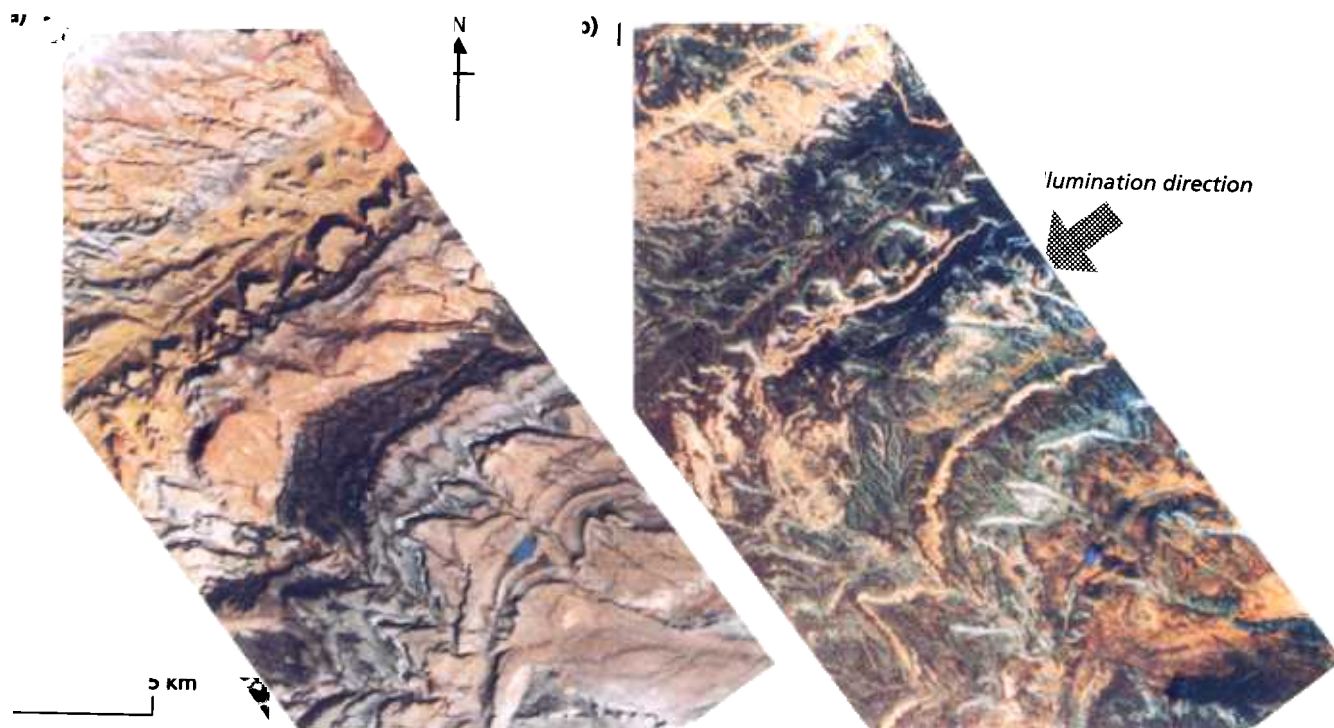


Plate 7.1 The images here cover a larger area of the Wind River Basin than do those in Fig. 7.24. Image (a) is a false-colour composite of Landsat TM bands 4, 3 and 2 as red, green and blue. The same primary colours in (b) are controlled by the HH, HV and VV polarized SAR data shown in Fig. 7.24. The colours in (b) therefore show the effects of surface morphology, back-scatter and of matter-radar interactions on the clarity of radar pulses. Those in (a) represent electronic and mineralogical transitions. Of particular interest are the red-hued areas in the south-east of (b), where strongly depolarized returns were produced. The lack of vegetation revealed by (a) in this area means that subsurface volume scattering may be responsible, possibly as a result of significantly deeper

weathering in the rock unit there. The other, more subtle, hues (pinks, creams, browns, greens and blues) relate more closely to variations in surface roughness. Compared with the TM image, the multipolarized radar image does not show such sharply defined geological boundaries and lithological distinctions. However, it does show additional boundaries and features that are not seen on any other kind of image, particularly in the north-west. An advantage of the control possible with SAR is the suppression of shadows resulting from illumination parallel to the main ridges in the area. Both images clearly complement each other and Plate 6.5. Courtesy of Tom Farr, Jet Propulsion Laboratory, Pasadena.



Plate 7.2 SIR-C/XSAR radar image of Erta A'le volcano in Ethiopia (see Fig. 7.25) using L-band as red, C-band as green and X-band as blue. The coloration helps distinguish lava flows of different ages and with different surface textures. The bluish flows are 'rough' only in X-band and may be pahoehoe flows, perhaps of obsidian, which is common on Erta A'le. Brownish flows are relatively smooth in all bands, and this, together with their being crossed by the bluish and pinkish flows, may indicate that they are old flows with surfaces that have been degraded by weathering and erosion. The white

areas face towards the radar illumination and are highlighted rather than being especially rough. Lakes are picked out as dark features. Much of the alluvial surround to the volcano is dark, suggesting that it has a smooth surface of silts and clays. The green areas at the top are of lush vegetation along drainages and spring lines at the base of the volcano. X-band radar penetrates the vegetation completely, whereas to L-band it is 'smooth', and only C-band wavelengths are back-scattered by leaves and twigs.

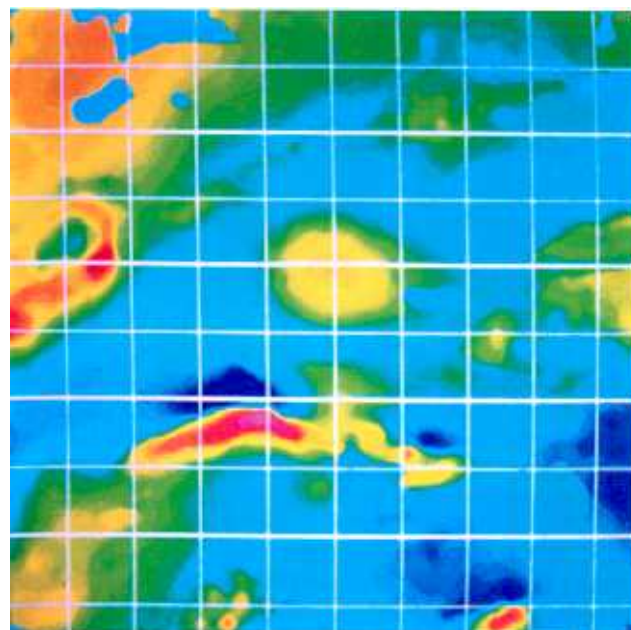


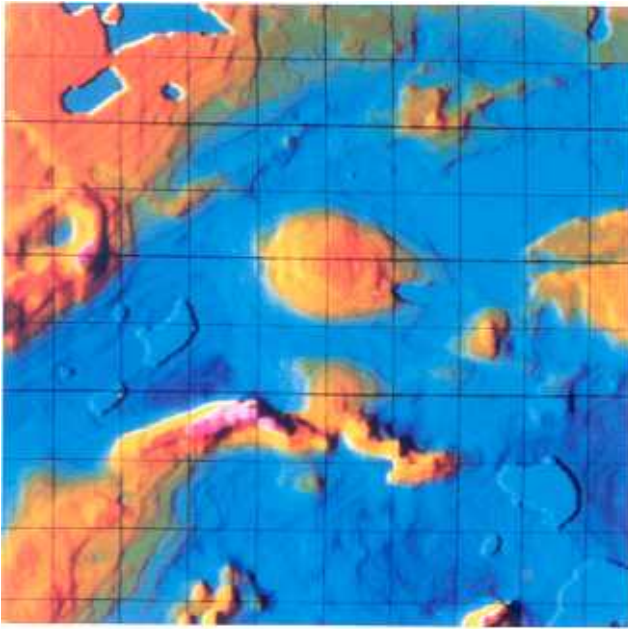
Plate 7.3 Recasting a frequency-filtered digital Seasat image in the intensity-hue-saturation (IHS) colour system restores the full information content of the image, but adds more easily interpreted colours relating to varying roughness and dielectric constant. In this case the Wamsutter Arch of Wyoming, USA, is more clearly defined by rock-related colour bands than in the original grey-tone image (Fig. 7.26a). Courtesy of Ronald Blom and Michael Daily, Jet Propulsion Laboratory, Pasadena.

Plate 7.4 The San Raphael Swell of Utah, USA, has been the target to test the geological usefulness of many remote-sensing systems (Plate 6.3 and Fig 6.12). In this image Seasat radar data have been combined with Landsat MSS data. The first, second and third principal components of Landsat MSS data have been modulated by the Seasat SAR data and displayed as red, green and blue. The extent to which the radar data have improved the image can be judged by comparing the four corners of the image, which Seasat did not cover, with the central parts. There are three distinct advantages with this method. The better resolution of the Seasat data resolution sharpens the expression of the Landsat data, topographic features are enhanced by radar, and variations in back-scatter resulting from surface properties control the brightness of the colours stemming from the decorrelated MSS data. A much more detailed geological interpretation is possible than would be the case with any single imaging system. Courtesy of Ronald Blom and Bill Stromberg of the Jet Propulsion Laboratory, Pasadena.



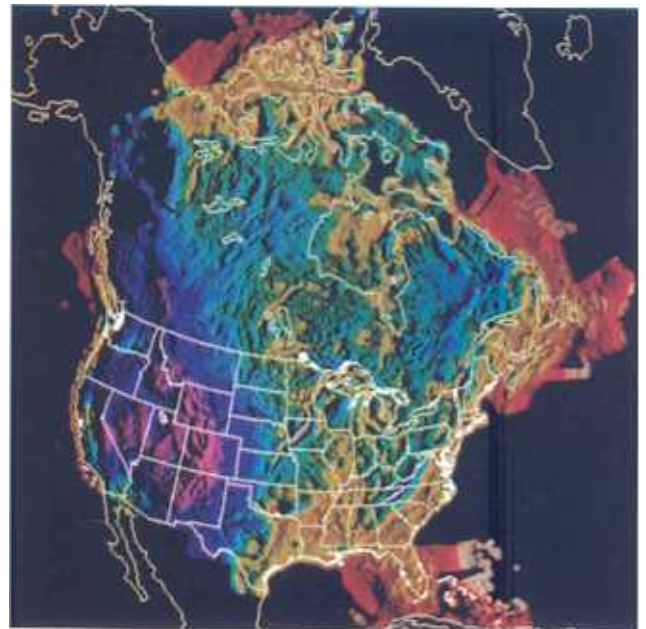
Plate 8.1 Version of Fig. 8.11(b) using look-up tables for red, green and blue to assign the 'rainbow' range from blue (low) to red (high) to the digital numbers (DN) of total magnetic field anomaly.





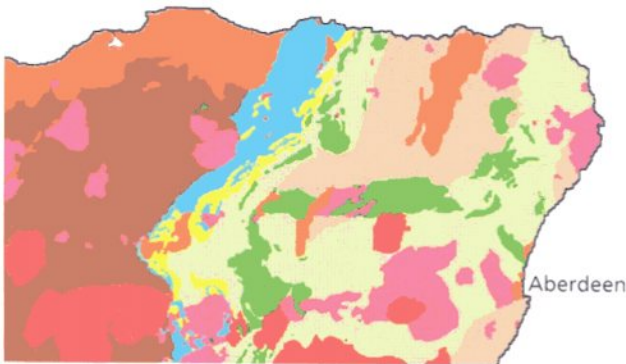
(a)

Plate 8.2 (a) 'Illuminated' aeromagnetic data from Fig. 8.13 combined with a coloured rendition. (b) Gravity anomaly data for North America colour coded from magenta through blue, green and yellow to red and 'illuminated' from the north-west. The large negative anomaly over the western USA is



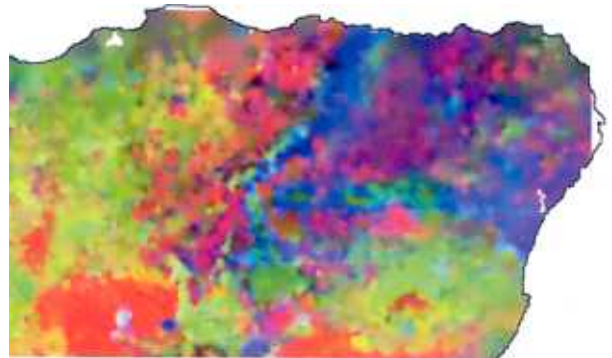
(b)

associated with active crustal thinning in the Basin and Range Province. Much detail of deep crustal structure is revealed for the central plains, which are veneered by almost featureless Mesozoic sediments. Courtesy of the Earth and Planetary Remote-Sensing Laboratory, Washington University, St Louis.



(a)

Plate 8.3 The geology of the north-east of Scotland (a) is among the most complex in Britain, comprising a metamorphic complex of Neoproterozoic sediments that have been repeatedly deformed by nappes and later regional folds. This complex also contains intrusions of granitic and basic-ultrabasic composition. The geological map is the result of over a century of field mapping in an area where exposure is at best 10–15%, and in some places is virtually nonexistent except along the coast. Moreover, it has been subject to repeated glaciation and most of the superficial sediments are exotic, albeit only a few metres thick. Image (b) is a rendition in RGB order of rubidium, strontium and chromium data interpolated from stream-sediment analyses. Sampling was in first-order streams, with catchments that are of the order of 1 km² or less (Fig. 8.7). Despite the exotic nature of most soils, the geochemical image bears a close resemblance to the mapped geology. This is because the streams have incised



(b)

through the superficial sediments as a result of active uplift following deglaciation, and sample bedrock directly. The red areas associate with granitic intrusions (red on a), being rich in Rb and low in both Sr and Cr, as expected from the general geochemistry of granites. Similarly, the blue areas—rich in Cr and poor in Sr and Rb—match the distribution of basic-ultrabasic intrusions (dark green on a). Green areas—high Sr, low Rb and Cr—occur in the dominantly metasedimentary areas (browns and pale green on a), although some of the high-Ca granitic masses (pink on a) are Sr-rich. The diffuse reddish area extending to top centre coincides with older migmatitic gneisses (blue on a). The coincidence of the geochemistry with known geology is sufficiently good that anomalies that do not relate to the geological map are likely to represent important unknown features, hidden by superficial sediments. Courtesy of the Director of the British Geological Survey.

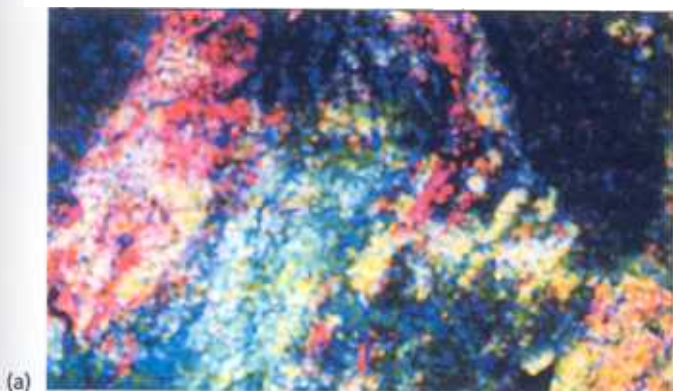


Plate 8.4 (a) Colour rendition of gamma-ray intensities for uranium, thorium and potassium as red, green and blue components for an area in Australia. Reds, greens and blues show areas that have high values for U, Th and K, respectively, and low values for the other two elements. Equally high values for all three elements are shown by grey tones and pastel shades. Dark areas have low radioactivity. Courtesy of BMR and CSIRO, Canberra. (b) Geological map of the corresponding area. Black lines indicate important faults and fold axes. Yellow areas are Quaternary unconsolidated sediments and all other shades relate to various Precambrian rocks. The brownish units at top right are glaciogenic sediments resting unconformably upon high-grade metamorphic rocks, the bulk of which are aluminous metapelites (pale purple), together with green amphibolites and various granitic rocks,



shown in blue, red and orange. Courtesy of the Geological Survey of New South Wales. Comparing the radiometric image with the map reveals some interesting features. Sharp boundaries on the image at top left and bottom right relate to major faults. The red to pink body on the image towards top left links to the red granite on the map, but extends over a larger area, perhaps indicating associated permeation of the metapelites by U-rich pegmatites. Some of the amphibolites show as dark streaks. The blue granite appears to be equally rich in all three radioactive elements, whereas the glaciogenic sediments have low radioactivity. The bulk of the metapelites appear to be Th-rich. Possibly more detail is revealed by the image than by the map. Courtesy of CSIRO, Mineral Physics Division, Australia.

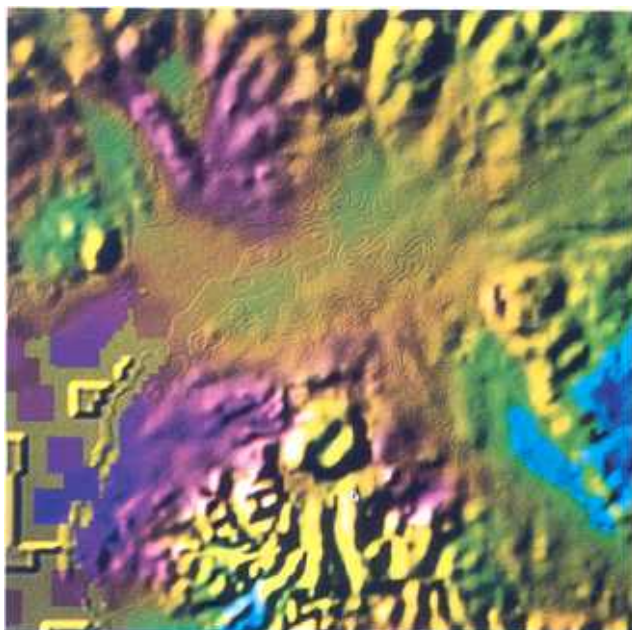


Plate 8.5 Image combining digital elevation data with Bouguer gravity anomaly data for the same area as Fig. 8.15, using the DEM illuminated from the north-west as intensity, a constant saturation and gravity as hue in an intensity-hue-saturation (IHS) transform. The colours range from blue for negative anomalies through a rainbow spectrum to magenta for the highest gravity anomaly. The green to yellow transition, being the most easily perceived spectral boundary, is keyed to the zero Bouguer anomaly.

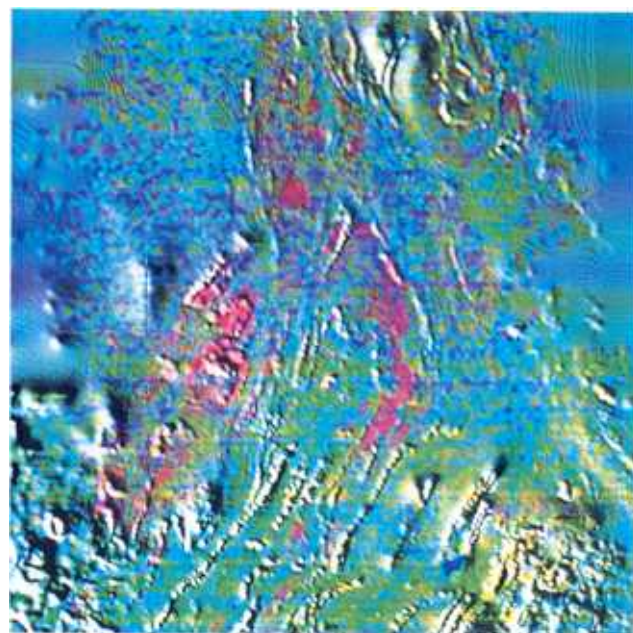


Plate 8.6 This image includes the area in Plate 8.4 and uses the intensity-hue-saturation (IHS) transform, with aeromagnetic data as intensity, a constant saturation and hue derived from U, Th and K data in Plate 8.4 (a). Although the faults are well displayed in both magnetic and radiometric data, the unconformity beneath glaciogenic sediments has no effect on the magnetic data, however, it is prominent in the radiometric data.



Plate 9.1 Enhanced Landsat TM image (bands 7, 4 and 2 as RGB) of an unmapped area in Eritrea showing clearly differentiated lithologies. The yellow and brown units are different marbles, red-magenta corresponds with metamorphosed basalts, and blue-grey units are chlorite and graphite schists. Bare and vegetated alluvium shows distinctly in the major drainage channels. The sharp contacts between different lithologies, and abundant examples of fine compositional banding help delineate major structures in the area. In the north-west quadrant, fine compositional banding (red, pale and bluish) has been shifted by strike-slip faults. The faulted rocks lie

within a tectonic unit that is bounded to the south-east by a sharp feature that cuts across the strike of the displaced banding to its north-west. This feature is a major shear zone. The yellow marbles that dominate the northern part of the area define several large folds. To the south-east of the marble zone are two thrust-bounded tectonic units, the thrusts being revealed by strike cut-offs. The lower of these units contains brown marbles and pale overlying rocks, whereas the upper unit is dominated by red metabasalts and blue schists. The complex is part of an accreted assembly of Neoproterozoic terranes that formed in a series of island arcs. Size: 20 x 25 km

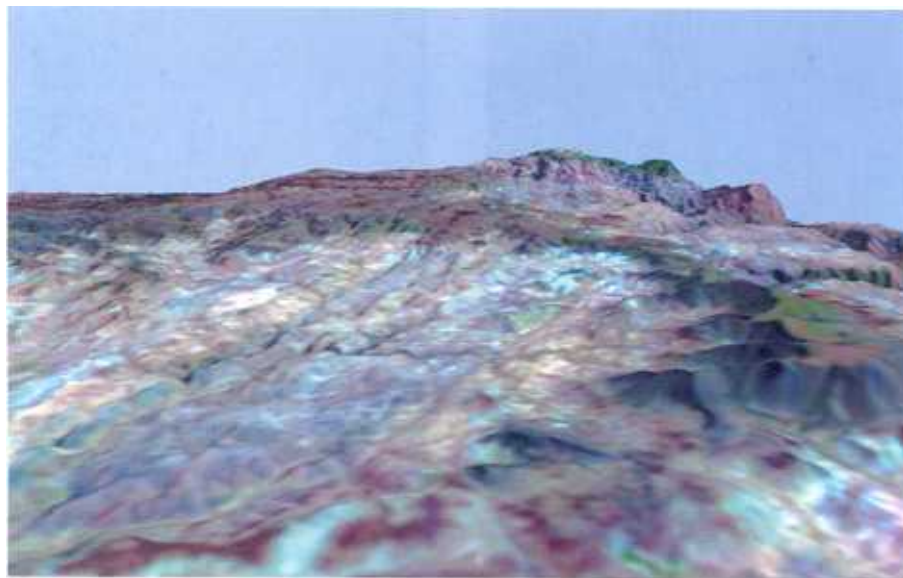


Plate 9.2 Perspective view combining enhanced Landsat bands 7, 4 and 2 as RGB with a digital elevation model (DEM) derived from stereoscopic aerial photography. This shows a dissected unconformity in north-east Africa that separates Tertiary continental flood basalts (dark grey with vegetation cover) from vertically dipping Proterozoic metasediments

with a strike that trends from lower left to upper right in the multihued terrain. The Tertiary cover forms a plateau at upper left, and a relict ridge at right. The horizon is about 15 km long. Exaggeration of the vertical scale gives a false impression of the slopes in the topography. The range of elevation in the area is from 1000 to 2400 m.

Plate 9.3 Image of part of Nova Scotia combining airborne SAR imagery with a geological map, using the intensity-hue-saturation (IHS) transform. The greens are Ordovician metasediments, reds indicate Devonian granites and the brown areas are mixed Carboniferous sediments. The radar adds considerable detail to the relatively uninformative geological map, particularly with respect to faults. Courtesy of J. Harris, Canadian Centre for Remote Sensing.

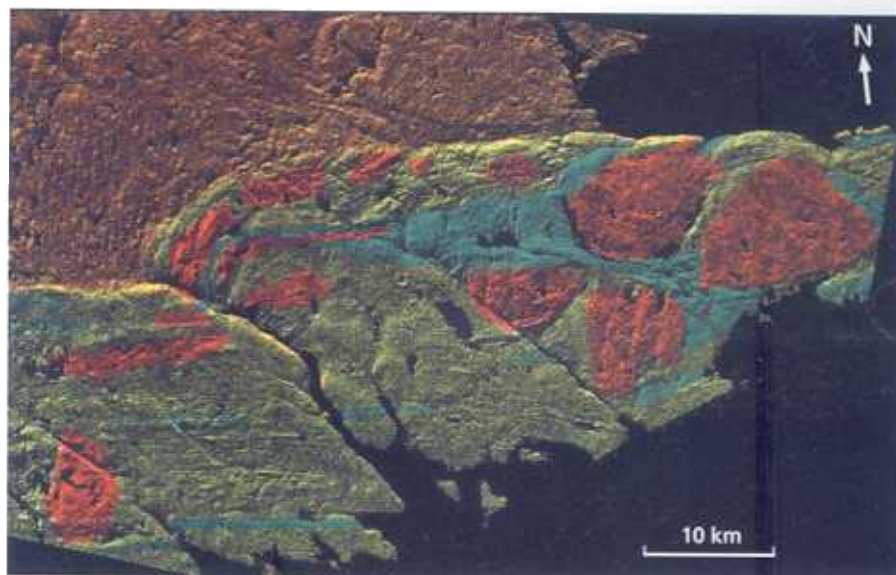


Plate 9.4 By combining SAR data as intensity and airborne gamma-ray spectrometer data (K, Th, U as RGB) as hue, an intensity-hue-saturation (IHS) transform reveals considerable structural and lithological variation for this area in Nova Scotia. Comparison with Plate 9.3 shows that the granites mapped in the field as similar bodies have considerable differences in U, Th and K content.

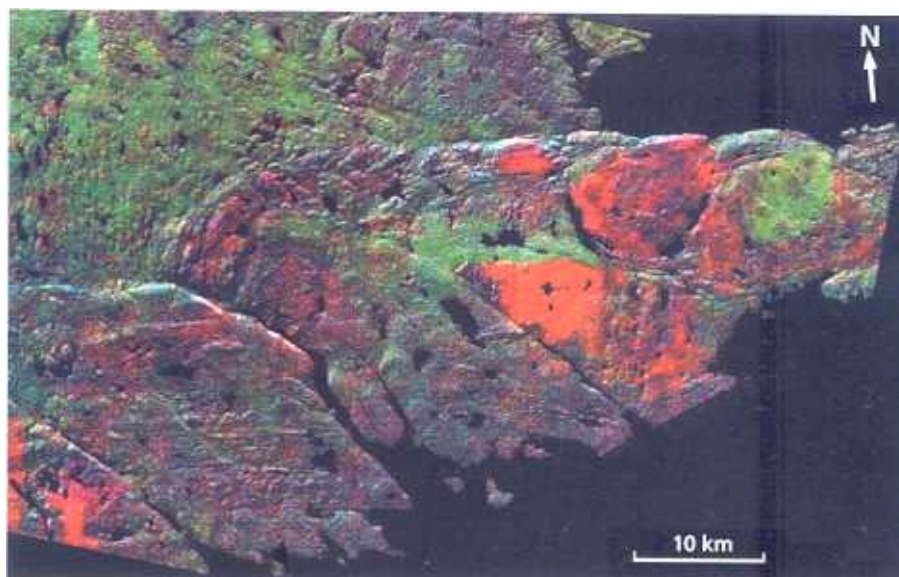
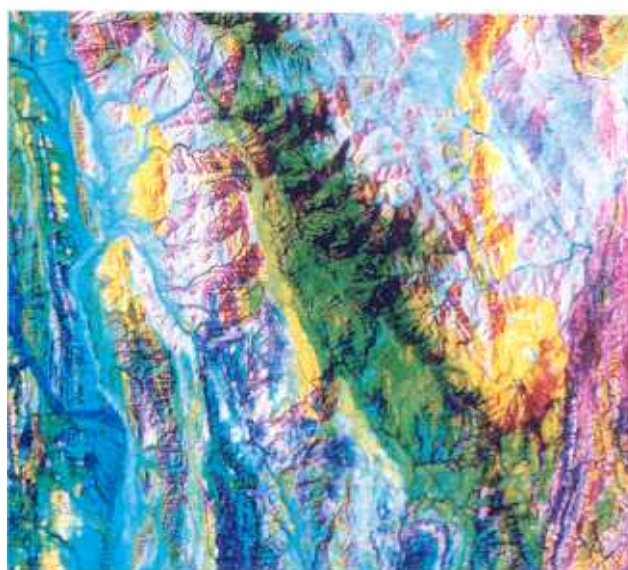


Plate 9.5 This false-colour ratio image of a geologically unknown area in Eritrea uses the ratios 5/4, 3/1 and 7/5 in RGB order to exploit the spectral features of the ferric iron- and hydroxyl-bearing minerals, which are often abundant in hydrothermally altered rocks, to assign a yellow colour to areas bearing them. The yellow areas on the left proved to coincide with hydrothermal alteration of metavolcanic rocks in which scheelite (CaWO_4) was found in quartz veins. The larger yellow zone running north-south is an area of gossanous material overlying a metamorphosed sulphide-rich shale, thereby showing that remote sensing has to be used with field verification. Size: 30 by 30 km.



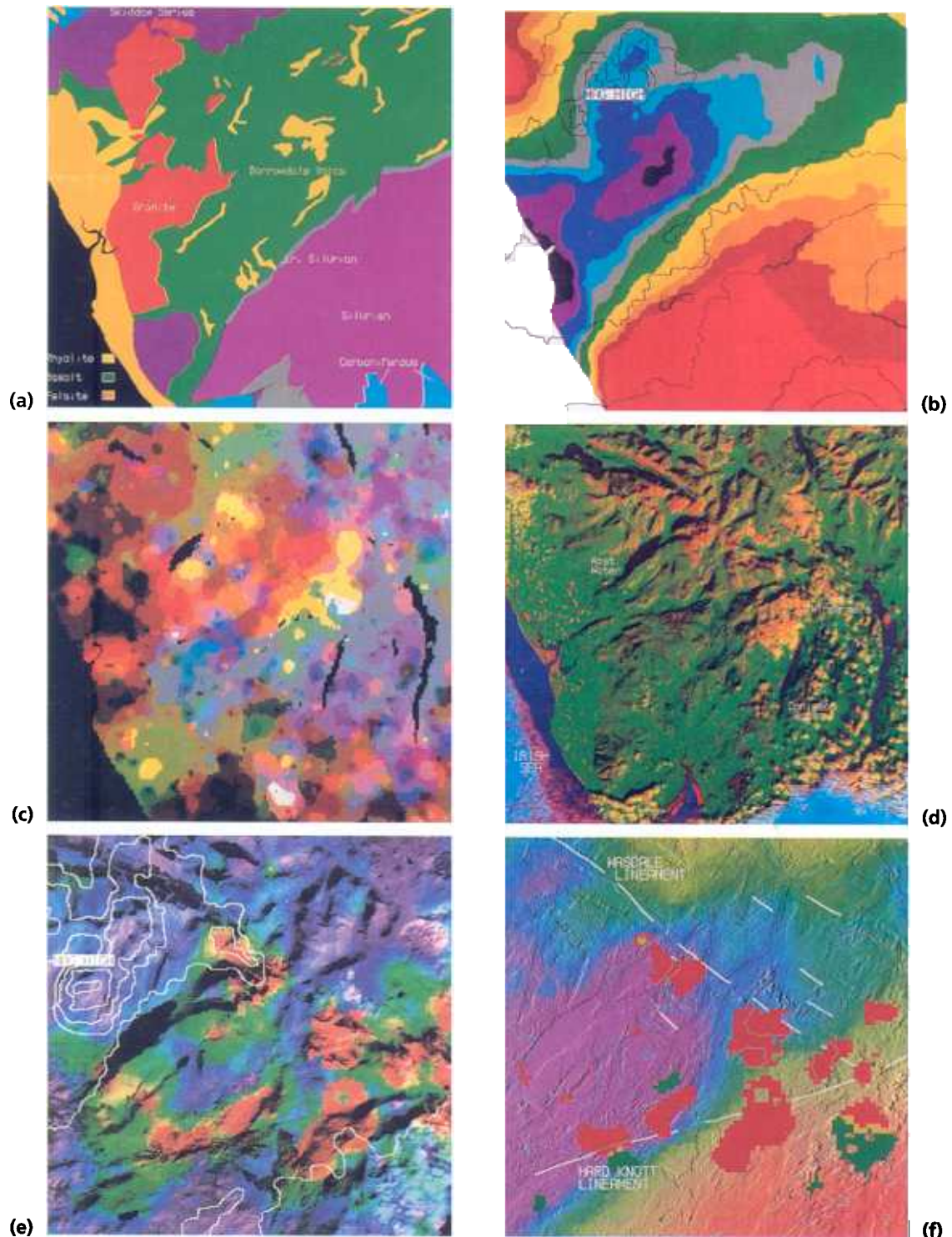


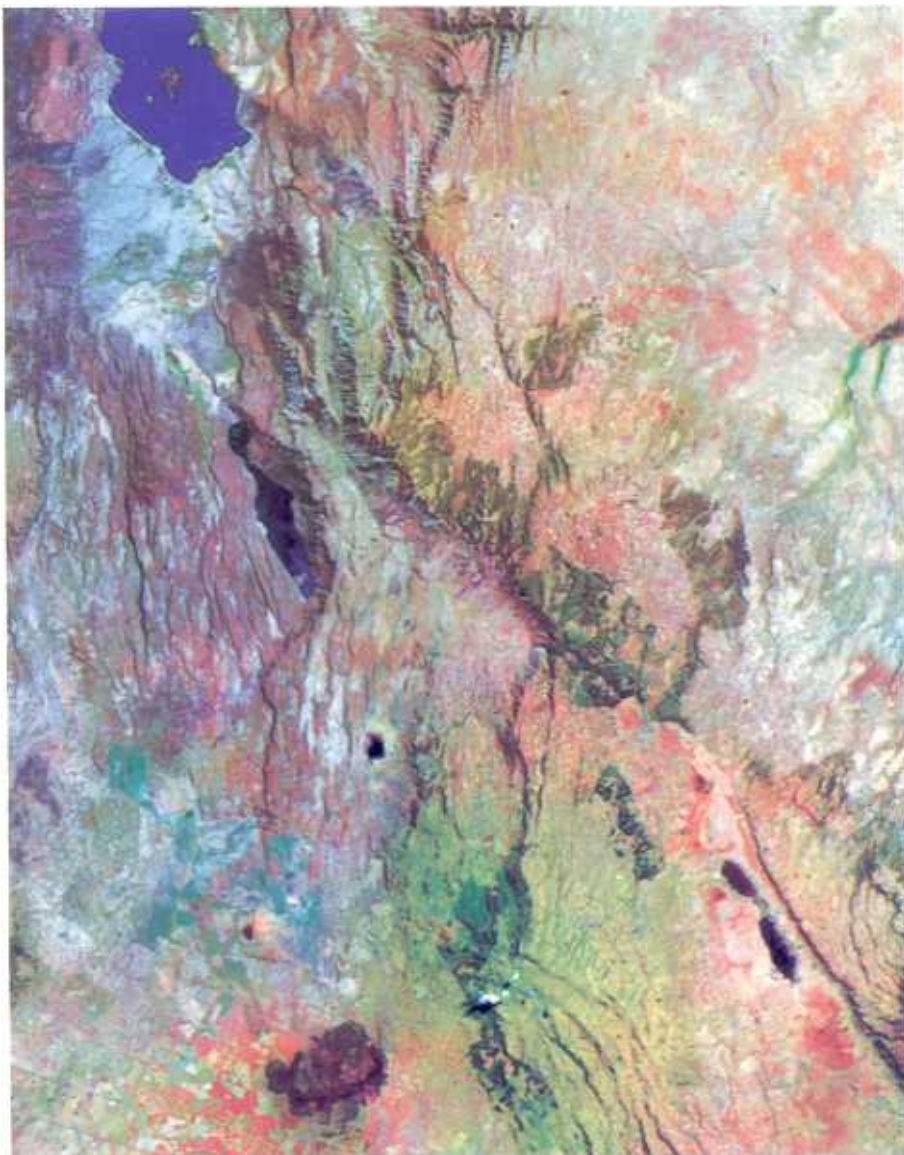
Plate 9.6 Data and results from an image-orientated geographical information system (GIS) exercise in targeting tin mineralization in the English Lake District. (a) Geological map: within the Borrowdale Volcanics greens are basalts and yellows rhyolites. (b) Colour sliced map of Bouguer gravity anomaly, ranging from magenta for lowest negative anomalies through blue to red at highest positive anomalies. Superimposed are contours for magnetic field intensity. (c) Tin, copper and zinc values sliced at 50–75–95–99 percentile levels (Fig. 8.8b) and displayed in RGB order. (d) Landsat TM 742 image showing the dominant vegetation

cover in green. (e) Landsat TM band 5 and tin values combined through the intensity–hue–saturation (IHS) transform, with magnetic field intensity shown in contours. (f) Directionally filtered Landsat TM band 5 image (to enhance fractures, the most important of which are in white) combined with Bouguer gravity anomaly data through the IHS transform. Superimposed in red and green are areas with significant tin and zinc–bismuth anomalies, respectively. The last two are spatially associated with fractures at the margin of the negative gravity anomaly above a buried granite. Courtesy of the Director of the British Geological Survey.

Plate 9.7 Stand of prematurely senescent maple trees (orange) related to seepage of natural gas. Courtesy of Harold Lang, Jet Propulsion Laboratory, Pasadena



Plate 9.8 Landsat TM 742 image of part of the East African Rift in Kenya, with vegetation related to springs near faults showing green in an otherwise arid terrain. Width 40 km. Courtesy of Eosat.



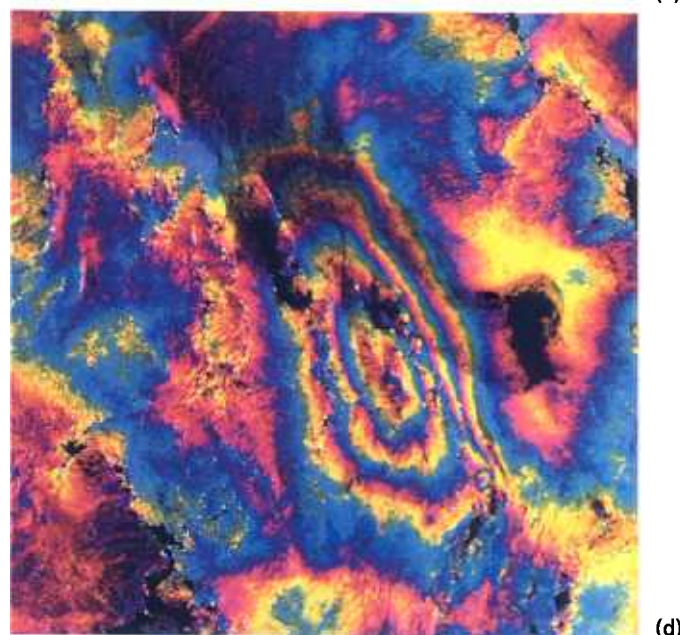
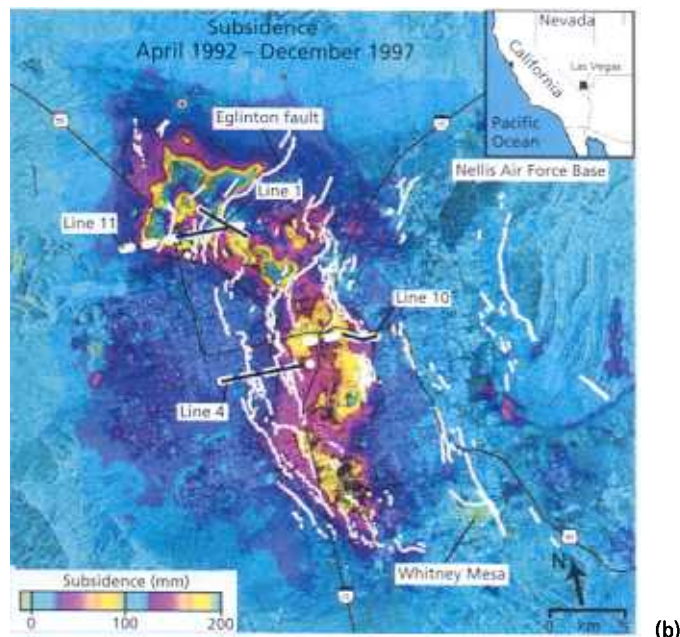
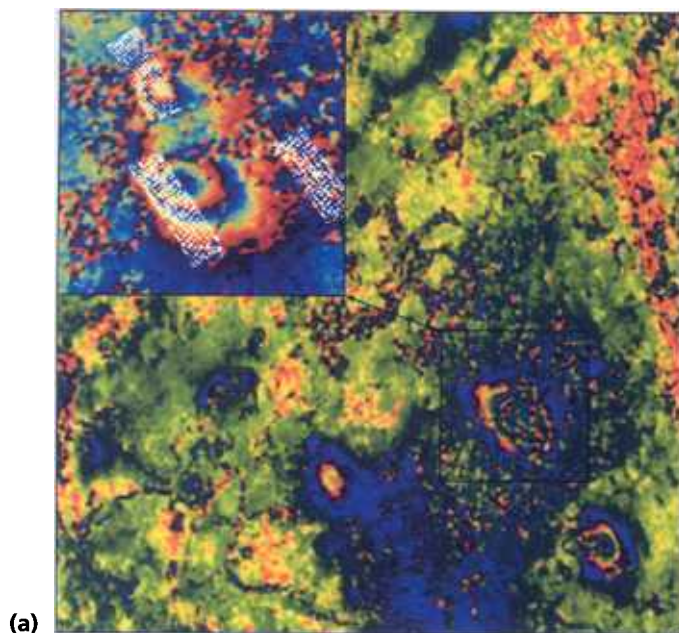


Plate 9.9 Examples of interferometric radar images that show changed ground elevations. (a) Interferogram of part of the Selby coalfield in northern England, produced from ERS SAR data separated by 35 days. The blue to red interference fringes each signify 28 mm of settlement, the total in this area being about 90 mm in the period covered. The inset is for a different period, and shows details of both subsidence and mining activities. The ornamented rectangles show the extents of subsurface workings in the coalfield. Courtesy of Trish Wright, Marconi Research Centre: ©Marconi Electronic Systems. (b) Interferogram of the Las Vegas area, Nevada, produced from ERS SAR data from April 1992 and December 1997. The interference fringes (see bottom left) each span 100 mm of settlement, giving a total over 5.8 years of 20 cm. In this case,

the settlement is due to abstraction of groundwater from beneath the city. Courtesy of Falk Amelung, Stanford University. (c) Interferogram superimposed over a perspective view of Mount Etna, in which several 28 mm fringes indicate the changes in surface elevation resulting from magma movement. Courtesy of Didier Massonnet, CNES, Toulouse. (d) 28 mm colour fringes related to ground subsidence caused by the 1993 Eureka Valley, California, earthquake (magnitude 6.1). The elongate basin subsided by around 10 cm during and after the earthquake. The dark areas are to the product of incoherent returns in the radar data, whereas the coloured areas beyond the seismic basin reflect shifts of a few millimetres, possibly as a result of soil swelling after rains. Courtesy of Jet Propulsion Laboratory, Pasadena.

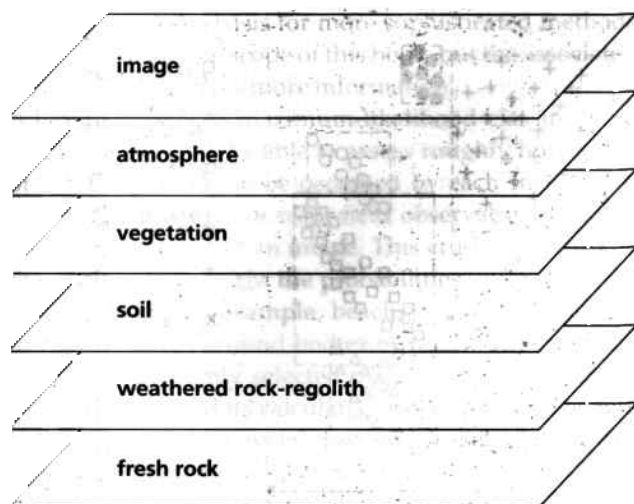


Fig. 5.31 Diagrammatic representation of the various possible layers that obscure or cryptify true geological relationships. To this may be added the limited number of mineral types that easily acquired remotely sensed data can distinguish. Even if completely bare of cover, rocks can rarely be classified on standard petrographic grounds by remote sensing alone.

cover. More often than not so many classes would be required that the results are so confusing as not to repay the effort expended. The only option is to visually interpret enhanced images of different kinds on the traditional basis of colour, texture, pattern and geological context. In the majority of applications this is the dominant approach, aided only where appropriate by judicious use of classification. Despite this pessimistic view, it is possible to detect and outline spectrally unusual rocks by automated means—many are interesting because their oddness stems from them being economically interesting concentrations. For that reason, briefly introducing some classification techniques is appropriate for the image-orientated geologist.

Inspection of the histograms for different surface categories allows different ranges of DN in the overall histogram to be assigned to those categories or groups of categories that can be separated by this means. Assigning a colour to each of these ranges highlights each pixel with a DN that falls within the range in the appropriate colour. This simplification of the data, in effect a crude classification, is called density slicing (Plate 5.8). Where only a few, distinct surface types make up an area, this can be a rapid and accurate means of mapping their spatial distribution. A more frequent use is to simplify a scene to assess gross variations, or to express continuous variation in a single type of surface. One example of the last application is density slicing of the VNIR/red ratio to express variations in density of vegetation cover. Another is slicing the DN range of blue or green reflected radiance from the bed of a body of clear water to express depth variations (Section 1.3.4, Plate 1.1).

Figures 5.28 and 5.29 show that some surface categories display contrasted reflectance at different wavelengths. Using two bands of data, selected because of such contrasts, can enhance differences between categories and improve their separability. This is shown diagrammatically in Fig. 5.32(a). To train a computer to make these separations means giving it a set of rules on which to base 'yes' or 'no' decisions. The simplest means of doing this is to divide the bivariate distribution into rectangular boxes (Fig. 5.32b). The boundaries of the boxes represent the ranges of DN for the two bands within small, known areas of the interesting surface categories. Such areas are selected during field work and are termed training areas, because they are used to train the computer. The computer compares the DN of unknown pixels with the boxes. If they fall within one they are assigned to the relevant class. If they fall in none then they remain unclassified, and possibly indicate a need to refine the means of classifying or increase the number of classes. The same principle can be applied to any number of dimensions, and usually is. This is known as parallelepiped classification. An important limitation is that natural categories generally plot as ellipsoids in N -dimensional space, and parallelepipeds are only a crude representation of this. A refinement is to represent the volume occupied by the data from training areas by an assembly of smaller parallelepipeds (Fig. 5.32c).

Providing the computer with a more sophisticated set of rules for decisions of the 'yes, if' type means using the statistics of data from training areas. The simplest of these classification routines is to identify the mean DN for each training-area category in each band being used. The distance from each of these means in N -dimensional space of an unknown pixel can then be calculated. The pixel is assigned to the class with the closest mean. Figure 5.32(d) shows this minimum-distance-to-means classification for the two-dimensional case. A refinement of this is to bring the variance of DN within training areas into play. Assuming that the DN within a training area form normal distributions for each band, the histograms can be considered to be bell-shaped. Depending on the variance (a measure of the width of the distribution) the further a DN is from the mean the less the probability that it represents the category in question. In the two-dimensional case (Fig. 5.32e) DN from a training area define an elliptical cluster. Each cluster can be contoured to show how the probability decreases away from the mean point. Figure 5.32(f) expresses this more graphically as a 'topography' of probability. The computer can then assess the plot of DN from an unknown pixel in this probabilistic context. It calculates the likelihood of the pixel belonging to each of the predefined classes and assigns it to that class where the likelihood is maximized. This method is maximum-likelihood classification. The three multispectral classification methods shown in

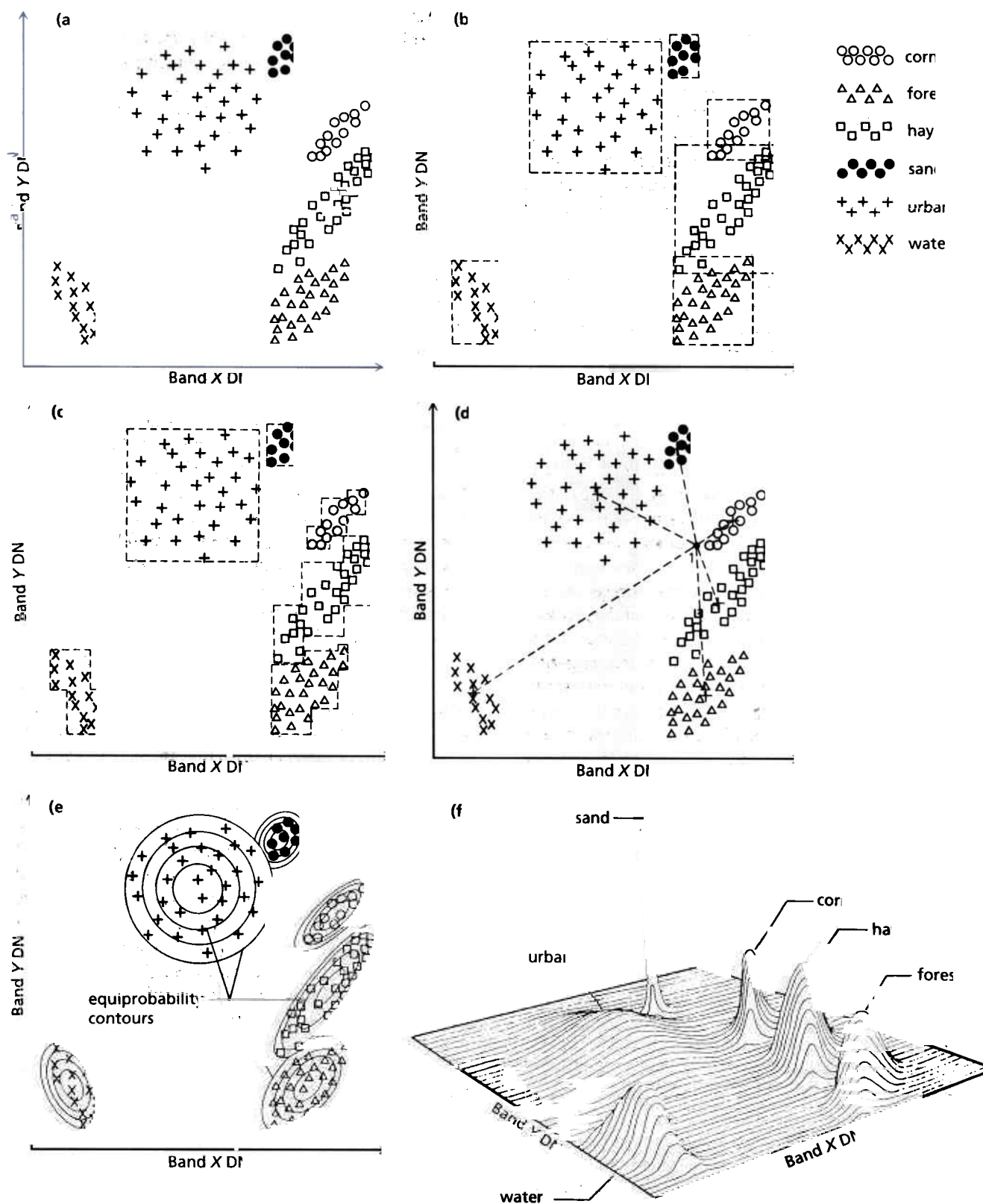


Fig. 5.32 Different methods of spectral classification can be represented diagrammatically by reference to bivariate plots of two of the bands used. In (a) several clusters in the data correspond to areas with different kinds of surface. The simple parallelpiped classification is shown in (b), that based on more precise boundaries for each class in (c), and (d)

represents the minimum-distance-to-means method. The contours in (e) express the probability that any point belongs to a particular class, which is the basis of the maximum-likelihood method of classification. The probabilities are illustrated more graphically as a three-dimensional topology in (f).

Fig. 5.32 form the basis for more sophisticated methods that are beyond the scope of this book, but the associated further reading gives more information.

Before making a maximum-likelihood classification, an interpreter may be able to assess roughly how much of a scene is likely to be occupied by each surface category. This may result from field observations or from a rapid inspection of an image. This crude estimate can then be used to weight the probabilities involved in the classification. For example, beach sand will be limited to narrow strips around bodies of water. The probability of any randomly selected pixel being beach sand is therefore low, and its calculated probability in that context is given a low weighting. Conversely, vegetation in a humid climate is very common, so the probability of any pixel being classified as such is given a high weighting. Another anticipated factor is that not all real surface categories are represented in the training areas. The classification can be weighted accordingly so that the computer does not try to force 100% of the scene into classes. Both these refinements help to ensure that the classification is as close to reality as permitted by the quality of the data and the precision of the classification method.

All the above classification methods rely on the operator making decisions about the areas on the grounds that are most representative of the surface categories of interest. This approach is termed supervised classification. It is also possible to allow the computer to examine the data in each band and sort out particular correlations amongst them. This in effect means dividing the N -dimensional histogram into arbitrary segments purely on the basis of heterogeneities in the distribution of DN. The operator can control the sharpness and number of the divisions, but basically leaves the computer to its own devices in assigning pixels to different classes. This is unsupervised classification, and expresses differences purely on the basis of spectral properties, i.e. it is empirical. It is useful in totally unknown areas, when it provides the operator with a means of identifying spectrally different kinds of terrain. This can then orientate the future field work, or be used as an adjunct to more conventional visual interpretation of the image.

Usually, the more independent variables, or dimensions, used in a classification the more easily spectrally distinct classes can be separated. However, in the 0.4–2.5 μm range there is a high degree of correlation between bands. This is quite clear from the spectral reflectance curves in Fig. 5.29, particularly those for rocks. Differences are expressed in only two or three bands—in the visible range expressing colour, that around 0.8–0.9 μm with an absorption feature related to oxidized iron minerals and that around 2.2 μm for the OH^- vibrational transition. These bands are the main means of discriminating iron-rich and clay-rich rocks from others.

By themselves these bands would perform a rough and ready classification. It might seem that using all the bands available from a remote-sensing system would sharpen the discrimination even further. However, the effect of strong interband correlation is to dilute the contribution of the most discriminating data. What happens in practise in this part of the EM spectrum is that as the number of bands used for classification increases beyond a certain limit the accuracy of the results actually declines. A strategy needs to be devised to use only those bands that do the job well. This approach also speeds up the operation of the software involved.

An irritating problem with classification based on spectral data alone is the variation in DN resulting from varying illumination. In terrain of high relief or where Sun elevation is low, the presence of many shadows disrupts the process. Moreover, the spectra of rocks contain only subtle features, which can be hidden in the natural variations. Variation in illumination is reduced by ratioing. Given carefully chosen bands, ratios can enhance these subtleties (Section 5.5). Another advantage of ratio techniques is that they combine data from several bands in a smaller number of parameters, thereby improving classification. A disadvantage is that they have a low signal-to-noise ratio, which tends to counteract some of their advantages, with noise pixels being misclassified.

Figure 5.32(e) shows that four of the six categories overlap completely in the bands defining the horizontal and vertical axes, even though the overall separation looks good. Unfortunately computers do not have eyes and work systematically with Cartesian co-ordinates in this case. A classification based on Fig. 5.32(e) would not be very good. If, however, the data could somehow be plotted on other axes, so that differences are maximized, that would improve the classification based on two bands alone. A more powerful approach is to perform a PCA analysis before classification. This not only derives new axes, but also expands the data to fill N -dimensional space as fully as possible. More important still, it combines together the highly correlated parts of the N -dimensional distribution along the different axes (Section 5.4). This reduces automatically the number of axes containing significant data. Each axis has a contribution from all the bands and the new data sets are uncorrelated. Further improvements deliberately attempt to arrange the new axes to maximize separation between the fields of known or suspected classes using directed PCA (Section 5.4.1), as in Fig. 5.33.

Classification for geological applications has two practical approaches. One is to map surface classes that correspond to several different kinds of underlying rock. This can be useful in producing a reconnaissance lithofacies map of a whole area, to be refined by field checking. In other words, the common rocks are sought. Plate 5.9 gives an example of this approach. Another

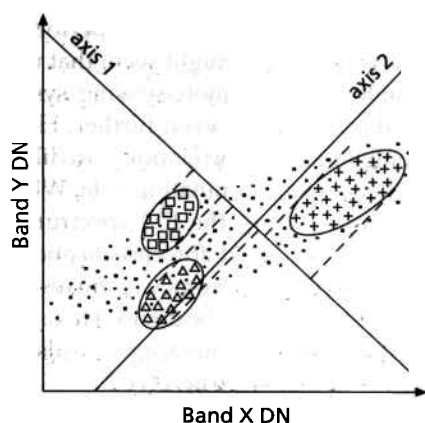


Fig. 5.33 Lying within the highly correlated ellipse of data on this bivariate plot of digital numbers (DN) for two bands are three classes of surface that need to be separated. It is clear that classification based on the two axes shown would produce confusion between the classes. However, by basing a principal component (PC) analysis on the means and covariances of data from within training areas of these classes produces new canonical axes, which maximize the differences between them.

strategy is classification to pick out those small, usually imperceptible areas on an image that contain rare but distinctive rock types. This is of great importance in commercial mineral exploration. First, ore deposits are unusual rarities by definition. Second, to find them by painstaking field survey is slow and expensive, especially in poorly known areas. If classification of remotely sensed data, even if it is quite imprecise, can narrow the area of exploration, so much the better. Field geologists, geochemists and geophysicists are then able to concentrate their talents much more cost-effectively.

Whether or not a particular patch of a surface category can be classified accurately depends on its size relative to the spatial resolution of the imaging system. It also depends on its contrast to the surrounding surface. In many cases an individual pixel's DN is contributed to by several surface components with sizes that are less than that of the pixel itself (Fig. 3.10). The DN of such a mixed pixel is an average of those of each of its components weighted by their relative size and their contrast. Strongly contrasted components may have such an effect on the pixel that their contributions to its DN are disproportionately large compared with their actual size. Under certain conditions very small isolated areas therefore can be detected and classified.

5.6.2 Spectral mapping

By collecting image data in many narrow, adjoining wavebands, hyperspectral imaging devices opened up a new dimension in remote sensing. As well as vastly increasing the number of possibly discriminatory combinations of bands, as RGB images, ratios and so on,

the scene corresponds to spectra of radiances for each pixel that makes it up. Consequently, these spectra can be compared with laboratory or field spectra of pure surface materials in ways that make possible identification of the mineralogical composition of natural surfaces. Before you become over-excited at the seeming possibility of geological mapping without the rigours of field work, remember that reflectance spectra address only the top few micrometres of the surface. That is, the spectra are dominated largely by minerals that form when rock-forming minerals are weathered, and they are not reflections of the petrographic and textural criteria that geologists use for rock classification. Moreover, as Fig. 5.31 shows, over most parts of the Earth rocks are obscured by soils (often transported) and various types and densities of vegetation. As if that were not depressing enough, there is a further complication; rocks, even when in a pristine state, are usually made up of mixtures of several prominent minerals. Nevertheless, hyperspectral data are so rich in information that they can more than while away those rainy Friday afternoons in the office!

A typical set of hyperspectral data involves 224 bands (AVIRIS), which means factorial $224 (224 \times 223 \times 222 \times \dots)$ possible RGB false colour images. If I am not mistaken, that is somewhat in excess of the number of hydrogen atoms in the known universe. Clearly, exploring the possibilities is not to be undertaken lightly; indeed for most combinations it is fruitless because of the high correlation between nearby narrow bands. The key to such an approach, which is useful because it maintains the data in visually appreciated form, is to remember that the sequence of bands builds-up for each pixel a reflectance spectrum that is almost as precise as one collected by a laboratory spectroradiometer. Knowledge of the wavelengths associated with various mineralogical transitions—absorption features, such as those associated with Fe^{3+} , Al-OH and Mg-OH , and C-O —allows bands that occur in and away from common features to be displayed as RGB combinations. They will show areas of surface that contain an abundance of the chosen mineral in vivid colours, whereas areas containing little will be dull greys. For an unknown area, that would be a good opening strategy. Some hyperspectral software packages show transects across a scene as stacked, one-dimensional RGB combinations of spectrally associated bands in increasingly long-wavelength regions of the data, allowing bands to be skipped to avoid the problem of redundancy. Where there are areas along the transect that contain high abundances of minerals with prominent features, the RGB 'slices' in different parts of the spectrum show as bright colours. Various means of enhancing contrast (Section 5.2) allow muted colours to be made more vivid and interpretable in these generally strongly correlated bands. Spatial filtering to

sharpen boundaries is less useful, as the narrow bandwidths mean that signal-to-noise ratios (Sections 3.3, 3.4) are considerably lower than in broader band data—edge enhancement makes the noise more obtrusive, whereas noise-removal filters may remove spectral features of small areas with unusual mineralogy.

The large number of bands, together with the high redundancy among most of them, suggests that data reduction by principal component analysis is an excellent strategy, both to save time and to highlight interesting areas and materials. As you saw in Table 5.1, the key to understanding the results of PC analysis is the eigenvector matrix. That shows the loading of different bands in each PC. In the case of bands that encompass different materials' absorption features, PCs that are strongly loaded by them express most clearly the distribution of those materials. For hyperspectral data, the eigenvector matrix is very large—about four pages of this book—and therefore difficult and confusing to use. An alternative strategy is to express the eigenvectors for each band as a curve for each PC. As the bands are arranged in increasing wavelength, this rendition can be compared with reference spectra of different materials. In this way the operator can select the PCs most affected by distinctive features in them. Similarly, the eigenvalues, or variances for the PCs can be plotted graphically. Yet it is still tedious to plough through as many as 224 plots—for high-order PCs may contain information on rare and spectrally odd materials as well as noise. The directed PC analysis approach helps overcome that, by performing the transform for mineralogically selected subsets of the data (Section 5.4.1). The division of Landsat TM bands, and those of other broad-band sensors into 'Fe³⁺' and 'OH' regions (visible and VNIR, and SWIR, respectively) for directed PC analysis expands with hyperspectral data to groups of bands that encompass, Fe³⁺, Al-OH, Mg-OH and C-O regions, and in a few cases specific minerals, such as alunite.

Using ratios between bands that express specific absorption features (Section 5.5) is also a possible strategy with hyperspectral data, being possible for several specific mineral groups. As well as reducing redundancy and helping focus on specific minerals for visual image interpretation, PC analysis and ratioing of hyperspectral data also assist in their digital classification, by reducing the number of dimensions involved.

As field identification of many of the minerals that show prominent spectral features is difficult, particularly when they are mixed, supervised classification of hyperspectral data is difficult without field or reference spectra. In their absence, there are several variants of empirical, or unsupervised classification. Data reduction by PC analysis or ratioing aims at the techniques for unsupervised classification outlined in Section 5.6.1, with ratios having the advantage of reducing spurious

classification according to variable shading of the surface. Other empirical methods use, in one way or another, the combined effect of hyperspectral bands to define spectra for each pixel in the scene.

Earlier in Chapter 5 you became familiar with bivariate plots of DN from broad-band data, where each pixel is expressed in two-dimensional space. These often show broad distinctions between different materials, as in Fig. 5.32. Hyperspectral data occupy so many bands that bivariate plots cannot even come close to expressing the full variability in a scene. For a computer, multidimensional or hyperspace is trivial, despite our difficulty in grasping it. The band values in a spectrum for a pixel plot as a point in hyperspace. Other pixels with similar spectral properties cluster in the same region of this n -dimensional space, whereas those with dissimilar spectra plot in distinctly different regions. The outcome is a series of 'clouds' in hyperspace. Two-dimensional 'slices' through the n -dimensional distribution would appear similar to the bivariate plots in Fig. 5.32, but clearly do not show all the 'clouds' that exist—more of this later. In this form, it might seem possible arbitrarily to define spectral 'clouds' or classes using any of the methods expressed by Fig. 5.32. However, this is inefficient. A faster method uses a process analogous to human brain functions—a neural-network technique (beyond the scope of this book). It starts by randomly selecting a number of positions in hyperspace, say 256, that arbitrarily represent the central co-ordinates of spectral classes. The position of each is joined by a line or vector through hyperspace to the origin of the distribution (0, 0, 0, . . . 0, 0). How similar each pixel's spectrum is to these classes is measured by the n -dimensional angle between its vector and that of the class centre. The first pixel assessed in this way will be closest to one of the arbitrary class centres, the position of which is adjusted to improve the fit. Progressively all pixels are considered, and the class centres evolve to positions with vectors that best fit all the spectra within the scene. Such an unsupervised, neural-network classifier is well described as a self-organizing mapper.

The same principle of exploring hyperspace lies behind methods of supervised classification, i.e. when there is some knowledge about the types of surface in an area. One begins with a training area that a known, mineralogically simple surface dominates, but includes other types of surface. The pixel spectra for the training area plotted in hyperspace will contain a cluster corresponding to the known material. By exploring plots of two-dimensional slices through the distribution, generally by a random rotation of the axes, that cluster can be isolated and defined in hyperspace. Its centre forms the basis for later classification. If field spectra are available, they too can be used to create n -dimensional centres. Using both methods of centre definition for classification depends

on how the distribution of spectra throughout the image are treated. One method focuses on calculating the angle between spectral vectors and those of the class centres. Low values indicate a close similarity of pixel spectra to the centres, and high values represent pixels that are unlikely to be of the same kind. Another assumes that all pixels contain varying proportions of the material that defines a centre, and therefore their spectra are mixtures of its spectrum and many unknown spectra. This method then calculates the proportion of the spectrum of each pixel that could be produced by incorporating that of the centre, a perfect match giving a result of 1.0 or 100%, and a complete absence of signs of the centre spectrum yielding 0. A third approach calculates correlation coefficients between each image spectrum and that of the centre, 1.0 showing a good match and 0 no match at all. These methods are not 'yes/no' classifiers, but produce continuous tone images that express probabilities. In that form their use is for assessing mineralogical variation in an area, but by using threshold values, distinct classes become possible. Variants combine results from several different centres as RGB images.

Field knowledge, either as spectra or training areas for surfaces dominated by specific minerals, can be used in another way. For example, let us assume that training areas or field spectra define all the most important materials (minerals and vegetation) that contribute to the surfaces in a scene. By assuming that all pixels contain spectral contributions from all the materials and their spectra result from linear mixing of these end-members, the linear unmixing method produces images corresponding to the calculated contributions of all the end-members. The sum of the fractional contributions in the perfect case will be 1.0 or 100% for all pixels. The critical stage in linear unmixing is ensuring that all possible end-members contribute to the calculation; a difficult undertaking even in well-known areas.

In a perfect world, geology would provide us with at least some outcrops containing absolutely pure minerals in which we are interested—that never happens! In a laboratory setting, however, such pure end-members can provide definitive spectra. Using laboratory spectra is the key to high-confidence spectral mapping, but hyperspectral images gathered from air- or spacecraft suffer from all manner of effects that are beyond our control. They measure *radiance*—the spectral radiant energy, originating in solar radiation, that is reflected from the Earth's surface. The spectral properties of the surface contribute to that, but so do other things. Solar energy varies in intensity with wavelength, partly because of the Stefan-Boltzmann and Wien laws (Section 1.2), partly from atmospheric scattering and absorption by gases in the air (Section 1.3.1), and also from the angle at which solar radiation strikes the surface (varying with sun angle and with topography).

Provided field spectra are obtained under similar illumination conditions to those at the time of imaging—preferably at the same time of the overpass—they are directly comparable with pixel spectra. Laboratory spectra suffer from none of these effects, being calibrated as *reflectance* relative to reference targets with flat spectra. Consequently, comparing image radiance and laboratory reflectance spectra depends on somehow calibrating the first to the second. This can be done by using standard models of atmospheric transmittance, but most conveniently by mimicking the way a laboratory spectroradiometer is used. Areas of quartz-rich sand or salt, and particularly concrete, have flat reflectance spectra (uniformly reflective at all wavelengths) and are highly reflective. If they are also topographically flat (as they often are), then an image spectrum of them will be an almost pure signature of atmospheric effects at the time the image was acquired, in the same manner as that of a laboratory spectroradiometer's reference target. Flat-field calibration divides each image spectrum by that of such a spectrally and topographically flat area, to produce a close approximation to a reflectance spectrum. Of course, in many parts of the world bare, flat expanses of any such spectrally monotonous materials are rare. An approximate calibration to comparability with laboratory reflectance spectra depends on estimating the atmospheric effect from all pixels in the scene. This involves calculating an average of all the image spectra, which, if the scene is sufficiently diverse, cancels out most of the features resulting from specific materials. Various adjustments prior to averaging, such as correcting for atmospheric scattering and instrumental gain by subtracting the average spectrum for the darkest pixels in the scene from all others, and adjusting all scene spectra so that the area beneath them is the same—effectively normalizing the brightness in every band—renders the average spectrum as a close approximation to that of a flat field. Each normalized spectrum divided by the average becomes an approximate reflectance spectrum.

Calibrated image spectra, being comparable with reflectance spectra of pure minerals obtained under laboratory conditions, potentially can be analysed quantitatively in terms of the mineralogical composition of the surface. Methods based on single class centres, defined by laboratory spectra, and on linear unmixing provide powerful tools for mapping. Many investigations when hyperspectral data become more widely and cheaply available will undoubtedly focus on exploration, that is on finding small occurrences of minerals that are good indicators of mineralization, such as various Fe^{3+} - and Al-OH-bearing products of hydrothermal alteration. Using laboratory spectra of anticipated indicator minerals, together with spectra of unmineralized, common surfaces, linear unmixing helps target areas on which ground exploration can focus.

Frequently, areas occur on various images that are quite clearly highly anomalous, yet their composition is unknown. With calibrated hyperspectral data, a spectrum from such an area can be matched against all reference spectra. Its degree of fit with standard minerals allows the composition of the anomalous surface to be assessed. There is a means of picking out such anomalies, however, without seeking them in the vast number of possible images that hyperspectral data present.

The exploration of image spectra in hyperspace by reference to arbitrary vectors though it forms the basis for another approach. Any pixel in a scene comprises various proportions of materials, either patches of pure materials, as analogized by the mixed pixels in Fig. 3.10, or intimate mixtures of say mineral grains or individual plants of different species. The spectrum of a pixel is the additive effect of those of all the materials that it contains, according to their proportions at the surface. Knowing the materials and their proportions would allow us to reconstruct the pixel's spectrum. Pixels composed of a single material have spectra that, by definition, are different from those of several mixed together—they are pure end-members, analogous to orthoclase, albite and anorthite in the feldspar solid solution series. So, in a hyperspace plot of image spectra, the spectra of pure end-members would be extreme values. The chances of finding pixels consisting of a pure materials decreases as pixel size increases, but if it was possible to find those that are extreme relative to the much more abundant mixed pixels, that would give a good idea of the range of materials that contribute to the scene as a whole. Once identified, their spectra can be matched to those of reference materials (see below). This gives a guide to analysis by the various techniques that I have outlined.

Estimating pixel purity relies on a computer assessing the position in hyperspace of each pixel spectrum relative to random vectors from the origin. Each analysis records the extreme pixels for a vector. As the process iterates through many such random vectors, it counts the number of times each pixel is an extreme. The number assigned eventually is an index of pixel purity, the higher it is the more likely a pixel is to consist of a pure end-member. Using colour density slicing (Section 5.6) helps highlight the most extreme pixels, and their spectra, matched against reference materials, suggest the regional end-members.

5.6.3 Spatial pattern recognition

The surface texture of different rock types, as expressed by drainage networks, colour banding and mottling for instance, is an important parameter in photointerpretation (Chapter 4). It represents the spatial distribution of tone variations. It predominates where tone, or DN, in an image varies with a high frequency. Where low

frequencies dominate, textural analysis is secondary in importance to evaluation of tone. As introduced in Chapter 2, human vision is very perceptive in recognizing textural features. However, it is very difficult for an interpreter to make accurate decisions on the boundaries between areas of subtly different texture—different interpreters would place the boundaries in different places. The use of computer-assisted measures of image textures, although very difficult and time-consuming, may add to the power of image classification. One method is to calculate the variance of DN within predetermined square boxes across an image. High variance suggests fine, sharp textures, whereas low variance indicates coarse, smooth textures. Another approach is to extract different spatial frequencies from an image using a variety of filters. Each expresses different kinds of texture, and the different images can form the basis for a semiquantitative classification of texture (Chapter 7).

One approach is to use statistical measures of the spatial frequency of variations in DN. To do this means that limits have to be set on gradients of DN and the spacing between changes in gradient must be represented in some way. This approach is likely to become increasingly useful in geological applications. Another line of attack is to use mathematical expressions of the spatial structure of the data. Basically, a sort of mathematical grammar is devised to describe each spatial pattern in the image data in terms of its parts. Each subpattern is broken down to its components, the smallest distinguishable category of which comprises primitives. Primitives may be the result of some kind of statistical approach to textural features. They are signified to the computer by a mathematical 'sentence', and each class of patterns is described by a set of such sentences. So far the use of pattern recognition of this kind has focused on simple or predictable patterns such as print characters in document image processing, chromosomes and military objects. Simple pattern recognition has been applied to geometric features in geology, such as the detection of circular and linear features. Although these are of considerable importance, the fact is that such features are found equally as easily by visual inspection (Chapter 2), and the automated methods often find spurious features.

Where a precise classification scheme exists for extremely complex patterns and textures, as in the case of fingerprints, considerable advances have been made. The problem with geological patterns is that, as well as being complex in their own right, they are also part of a complex whole. They have to be detected as well as being described and classified. Fingerprints are not detected by computers, they are sought, found and photographed by detectives. Elegant though automated spatial pattern recognition may be, it remains to be seen if it can be developed for new geological applications. More important is the question of whether it will repay

the effort in comparison with the biased, but extremely effective analysis by trained geological interpreters. One great difference between the military pattern analyst and the geologist is that the former generally is not permitted to visit the site of interest!

The spatial variation of DN in an image can be used in another way that might pay dividends for geological interpretation. That is by extending the methods of spectral pattern recognition using the concept of context. A particular pixel on a surface may be classified better if there is some knowledge of the classification of its surroundings. Contextual classification at its simplest consists of using the classification of other neighbouring pixels to weight the classification algorithm. The computer performs a check of results within a specified radius in order to achieve this. Conceivably it could be improved to take into account the context of other associated classes. Thus a pixel, which could be either granite or sandstone on the basis of its spectral properties, would be more likely to be sandstone if it were near to pixels identified unambiguously as limestone.

Any classification seeks to locate the boundaries between surface areas of different types and the extent of areas of each class as accurately as possible. A means of doing this is image segmentation. It consists of creating areas of several pixels that show spectral similarities. In a simple example a box of predetermined size passes over the image and tests the pixels falling within it for similarity according to the analyst's instructions. It is a technique that finds a ready use in analysis of imaging spectrometer data, such as those from AVIRIS, where the spectra of individual pixels are compared with spectral libraries of diagnostic minerals, such as different hydroxyl-bearing minerals (Plate 3.2). Either the spectra are matched directly or a kind of mathematical shorthand for the spectral shapes is devised to speed up the comparisons. If they fall within predetermined bounds of similarity they are grouped together as members of a class. The effect is to cause areas with similarity to spectral categories to grow to their predetermined spectral limits. By combining this approach with a means of edge detection (Section 5.3) the growth of areas of each class can be stopped at what are hoped to be natural boundaries in the scene. This approach is most successful in images where the surface categories are distinct and spectrally homogeneous, as in agricultural land. As emphasized in Section 5.6.1 geological classes are often spectrally heterogeneous.

5.6.4 Change detection

The monitoring of changes at the Earth's surface by the use of images recorded on several dates—multi-temporal analysis—has been important for studies of vegetation, cultural and military features throughout the history of

remote sensing. With photographic data it consists simply of visual comparison. The advent of regular repetitive digital imaging, first from meteorological satellites then the early Landsats, meant that more automated methods became possible. Most applications have been for analysis of the evolution of weather systems, crop rotation and diseases, changes in land-use patterns and many other topics for which these satellites were designed. One of the two main methods of analysis is based on detection of spectral changes in fixed surface materials. One example exploits the seasonal changes in vegetation canopy, which can discriminate between different species if the correct times for comparison are chosen. Another is detection of spatial change at the surface, such as the replacement of agricultural land by buildings, or the development of military installations. At its simplest the comparison is achieved by flicking from an image taken at one time to that taken at another.

Early in the Landsat programme it became clear that there was considerable potential for exploiting images taken at different times of year in geological investigations. As seasonal climates become more contrasted so lithologically controlled variations in soil moisture and vegetation become more variable during the year. Increasing soil moisture in bare areas results in lower reflectance in all bands, whereas increasing vegetation cover increases infrared reflectance and diminishes that in red and blue bands. The differences are most clear between dry- and wet-season images (Plate 5.10). For most purposes geologists exploit these differences in selecting an optimum time of year when geological information content is highest. In some climates the dry season is best because wet-season images are masked by vegetation, often with agricultural patterns. Others show geology better just after the wet season when soil and rock moisture varies most, according to porosity, and natural vegetation is strongly controlled by it.

Highlighting differences by using data from two or more seasons can be a very powerful technique. As well as expanding the variety of geologically related information it can show up geobotanical anomalies related to different kinds of geochemical stress and selective growth of plants. As the spectral effects may be at a maximum for only short periods, dates of imagery need to be chosen carefully. Sometimes particularly severe weather in one year is the only reason why such phenomena are detectable, and comparison is made between several years as well as several seasons.

A special case of multi-temporal analysis of spectral change is the use of diurnal variations in surface temperature to estimate the thermal inertia of the surface. This is considered separately in Chapter 6.

Short-lived geological and hydrological events can be detected, monitored and assessed by multi-temporal analysis. Among these spatial changes are volcanism,

faulting, slumping, floods, snow accumulation, glacial advance and retreat, shifts in drainage, sand bars, beaches and dunes, the progress of mining and the spread of subterranean fires. The data used are of the 'before and after' variety. With Landsat data the minimum frequency of imaging of an area of interest is 16–18 days. The pointability of the SPOT system increases this frequency to once every 4 days. For really vast events and those of short duration the TIROS-N/NOAA satellites produce coarse resolution data every 6 h.

The simplest means of automated change detection involves the subtraction of data for one day from that for another. This can present problems, however. Unless the two images have nearly the same structure—similar vegetation and atmospheric conditions—the new data sets can be overwhelmed by nongeological differences. It is best applied to data from the same season for different years. Another method depends on changes being poorly correlated with the rest of the data. They comprise anomalies on plots of DN from the same bands for different dates. Principal component analysis (Section 5.4) of selected bands for two or more dates casts the highly correlated data from the separate images into the first and second components. The major differences resulting purely from gross seasonal changes in vegetation appear in the second and third components. The unusual, small-scale differences are generally found in higher order components. Although the potential for becoming lost in a labyrinth of decorrelated data sets is high, careful selection of components and their display can highlight small but important changes very proficiently.

Associated resources on the CD-ROM

The image-processing techniques that I have outlined in Chapter 5 are covered by a series of practical exercises, which use a professional image-processing and digital-mapping package (TNTlite) on the CD-ROM. You can access the text files for these by opening the HTML file Menu.htm in the folder Imint on the CD using your Web browser. More detailed information about installing TNTlite is in Appendix C. As well as the data files (Imint\Exercises\Data) that the exercises use, you can import other data into TNTlite, provided that they are within the size limits that the vendors have imposed for commercial reasons. This means that designing new exercises is possible, and you can use TNTlite for small research projects.

My choice of TNTlite as a 'hands-on' resource is because MicroImages Inc (its designer and vendor) is the only software company that offers a size-limited, but fully functional version of its product as freeware for unlimited redistribution. Other than that, neither I nor Blackwell Science imply any endorsement of MicroImages products.

Since Landsat-1 made multispectral, digital imagery available publicly and cheaply in 1972, the potential for adding value to 'raw' remotely sensed data was enormous. There has been an explosion of software aimed at image enhancement and analysis over the last three decades. Some packages have disappeared as competition sharpened, whereas others have emerged as new operating systems made complex tasks more easily grasped by inexperienced users of computers. Competition, and the appearance of new kinds of data, such as digital SAR, hyperspectral imagery and DEMs, has meant that the functionality of these packages has evolved rapidly. At the time of writing, this growth seems to have reached a plateau. There has never been a more propitious time for entering the image processing and analysis fray, as the speed, precision and cost of computers and useful peripherals dwarf those when it all began to take off in the 1970s. Equally, the explosion in data sources, both commercial and freely available in the public domain, means that geologists risk being marginalized if they remain unskilled in this area.

The leading software vendors maintain lively Web sites, where you can browse their products. The market leaders are:

ERDAS—image and related software

<http://www.erdas.com>

ENVI

<http://www.rsinc.com>

ERMapper

<http://www.ermapper.com>

PCI

<http://www.pci.com>

Image Analyst

<http://www.nihon.intergraph.com/photogrammetry/rs/ianaly.asp>

TNTmips

<http://www.microimages.com>

An excellent review of these packages is in Limp, F.W. 1999. Image processing software: system selection depends on user needs. *GeoWorld* May, 36–46. Two monthly magazines provide breaking news of software and associated hardware developments—EOM (Earth Observation Magazine <http://www.eomonline.com>) and Mapping Awareness (<http://www.geoplace.com>).

Further reading

Section 5.6.2 Spectral mapping is condensed from TNTlite 'Getting Started' booklets *Introduction to Hyperspectral Imaging*, and *Analysing Hyperspectral Images*.

The following references form a framework for extending your understanding of digital image processing techniques.

- Anuta, P.E. (1977) Computer-assisted analysis techniques for remote sensing data interpretation. *Geophysics* **42**, 468–481.
- Berhe, S.M. & Rothery, D.A. (1986) Interactive processing of satellite images for structural and lithological mapping in north-east Africa. *Geological Magazine* **123**, 393–403.
- Blodget, H.W. & Brown, G.F. (1982) *Geological mapping by use of computer-enhanced imagery in western Saudi Arabia*. U.S. Geological Survey Professional Paper 1153.
- Buchanan, M.D. & Pendgrass, R. (1980) Digital image processing: can intensity, hue and saturation replace red, green and blue? *Electro-optical Systems Design* **12**, 29–36.
- Cañas, A.A.D. & Barnett, M.E. (1985) The generation and interpretation of false-colour principal component images. *International Journal of Remote Sensing* **6**, 867–881.
- Castleman, K.R. (1977) *Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey.
- Chavez, P.S. (1988) An improved dark object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment* **24**, 459–479.
- Chavez, P.S. (1989) Radiometric calibration of Landsat Thematic Mapper multispectral images. *Photogrammetric Engineering and Remote Sensing* **55**, 1285–1294.
- Chavez, P.S. & Bauer, B. (1982) An automatic optimum kernel-size selection technique for edge enhancement. *Remote Sensing of Environment* **12**, 23–38.
- Chavez, P.S., Berlin, G.L. & Sowers, L.B. (1982) Statistical method for selecting Landsat MSS ratios. *Journal of Applied Photographic Engineering* **8**, 23–30.
- Condit, C.D. & Chavez, P.S. (1979) Basic concepts of computerized digital image processing for geologists. U.S. Geological Survey Bulletin **1462**.
- Crippen, R.E. (1988) The dangers of underestimating the importance of data adjustment in band ratioing. *International Journal of Remote Sensing* **9**, 767–776.
- Davis, P.A. & Berlin, G.L. (1989) Rock discrimination in the complex geological environment of Jabal Salma, Saudi Arabia, using Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing* **55**, 1147–1160.
- De Souza Filho, C.R., Denniss, A.M., Drury, S.A., Rothery, D.A. & Carlton, R.W. (1996) Restoration of noise-corrupted optical Fuyo-1 (JERS-1) data using frequency domain techniques. *Photogrammetric Engineering and Remote Sensing* **62**, 1037–1047.
- Drury, S.A. (1985) Applications of digital image enhancement in regional tectonic mapping of South India. *Proceedings of the 18th International Symposium on Remote Sensing of Environment, Paris 1895–1904*.
- Drury, S.A. (1986) Remote sensing of geological structure in temperate agricultural terrains. *Geological Magazine* **123**, 113–121.
- Drury, S.A. & Hunt, G.A. (1988) Remote sensing of lateritized Archaean greenstone terrain: Marshall Pool area, north-eastern Yilgarn Block, Western Australia. *Photogrammetric Engineering and Remote Sensing* **54**, 1717–1725.
- Drury, S.A. & Hunt, G.A. (1989) Geological uses of remotely sensed reflected and emitted data of lateritized Archaean terrain in Western Australia. *International Journal of Remote Sensing* **10**, 475–497.
- Dykstra, J.D. & Birnie, R.W. (1979) Reconnaissance geological mapping in Chagai Hills, Baluchistan, Pakistan, by computer processing of Landsat data. *Bulletin, American Association of Petroleum Geologists* **63**, 1490–1503.
- Evans, D. (1988) Multisensor classification of sedimentary rocks. *Remote Sensing of Environment* **25**, 129–144.
- Evans, D. & Lang, H. (1985) Techniques for multisensor image analysis. *Proceedings of the 18th International Symposium on Remote Sensing of Environment, Paris*.
- Fabbri, A.G. (1984). *Image Processing of Geological Data*. Van Nostrand Reinhold, New York.
- Gillespie, A.R. (1980) Digital techniques of image enhancement. In *Remote Sensing in Geology* (eds B.S. Siegal & A.R. Gillespie), pp. 139–226. Wiley, New York.
- Gillespie, A.R., Kahle, A.B. & Walker, R.E. (1986) Colour enhancement of highly correlated images: I—decorrelation and HSI contrast stretches. *Remote Sensing of Environment* **20**, 209–235.
- Gillespie, A.R., Kahle, A.B. & Walker, R.E. (1986) Colour enhancement of highly correlated images: II—channel ratios and chromaticity transformation techniques. *Remote Sensing of Environment* **22**, 343–365.
- Gong, P. (1996) Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural-network techniques for geological mapping. *Photogrammetric Engineering and Remote Sensing* **62**, 512–523.
- Gonzalez, R.C. & Wintz, P. (1977) *Digital Image Processing*. Addison-Wesley, Reading, Massachusetts.
- Grasso, D.N. (1993) Application of the IHS colour transformation for 1 : 24 000-scale geological mapping: a low-cost SPOT alternative. *Photogrammetric Engineering and Remote Sensing* **59**, 73–80.
- Hardcastle, K.C. (1995) Photolineament factor: a new computer-aided method for remotely sensing the degree to which bedrock is fractured. *Photogrammetric Engineering and Remote Sensing* **61**, 739–747.
- Hord, R.M. (1982) *Digital Image Processing of Remotely Sensed Data*. Academic Press, New York.
- Hunt, G.A. (1988) Improvements in the forward and inverse principal component transformations for geological mapping in a semiarid terrain. *Proceedings, IGARSS Symposium 1988, Remote Sensing: Moving towards the 21st Century*, pp. 1061–1062. Special Publication 284, European Space Agency, Paris.
- Jensen, J.R. (1986) *Introductory Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey.
- Knepper, D.H. Jr & Raines, G.L. (1985) Determining stretch parameters for lithologic discrimination on Landsat MSS band ratio images. *Photogrammetric Engineering and Remote Sensing* **51**, 63–70.
- Lang, H.R., Adams, S.L., Conel, J.E., McGuffie, B.A., Paylor, E.D. & Walker, R.E. (1987) Multispectral remote sensing as a stratigraphic and structural tool, Wind River Basin and Big Horn River Basin areas, Wyoming. *Bulletin, American Association of Petroleum Geologists* **71**, 389–402.
- Mah, A., Taylor, G.R., Lennox, P. & Balla, L. (1995) Lineament analysis of Landsat Thematic Mapper images, Northern Territory, Australia. *Photogrammetric Engineering and Remote Sensing* **61**, 761–773.
- Mather, P.M. (1987) *Computer Processing of Remotely Sensed Images: an Introduction*. Wiley, New York.
- Mather, P.M., Tso, B. & Koch, M. (1998) An evaluation of Landsat TM spectral data and SAR-derived textural information for lithological discrimination in the Red Sea Hills, Sudan. *International Journal of Remote Sensing* **19**, 587–604.
- Moore, G.K. & Waltz, F.A. (1983) Objective procedures for lineament enhancement and extraction. *Photogrammetric Engineering and Remote Sensing* **49**, 641–647.

- Podwysocki, M.H., Segal, D.B. & Abrams, M.J. (1983) Use of multispectral scanner images for assessment of hydrothermal alteration in the Marysvale, Utah, mining area. *Economic Geology* **78**, 675–687.
- Pratt, W.K. (1978) *Digital Image Processing*. Wiley, New York.
- Renz, Andrew, N. (ed.) (1983) Remote Sensing for the Earth Sciences. *Manual of Remote Sensing*, 3rd edn, Vol. 3. Wiley, New York.
- Rothery, D.A. (1984) Reflectances of ophiolite rocks in the Landsat MSS bands: relevance to lithological mapping by remote sensing. *Journal of the Geological Society of London* **141**, 933–939.
- Rothery, D.A. (1985) Interactive processing of satellite images for geological interpretation. *Geological Magazine* **122**, 57–63.
- Rothery, D.A. & Francis, P.W. (1985) A remote sensing study of a sector collapse volcano. *Proceedings of the 18th International Symposium on Remote Sensing of Environment, Paris*.
- Rowan, L.C. & Bowers, T.L. (1995) Analysis of linear features mapped in Landsat Thematic Mapper and side-looking airborne radar images of the Reno 1° × 2° Quadrangle, Nevada and California: implications for mineral resource studies. *Photogrammetric Engineering and Remote Sensing* **61**, 749–773.
- Rowan, L.C., Goetz, A.F.H. & Abott, E. (1987) Analysis of Shuttle multispectral infrared radiometer measurement of the Western Saudi Arabian shield. *Geophysics* **52**, 907–923.
- Rubin, T.D. (1993) Spectral mapping with image enhancement is difficult. *Photogrammetric Engineering and Remote Sensing* **59**, 1191–1192.
- Ruiz-Armenta, J.R. & Prol-Ledesma, R.M. (1998) Testing Equations for enhancing the spectral response of hydrothermal minerals in Thematic Mapper images of Central Mexico. *International Journal of Remote Sensing* **19**, 1981–2000.
- Schowengerdt, R.A. (1983) *Techniques for Image Processing and Classification in Remote Sensing*. Academic Press, New York.
- Siegal, B.S. & Abrams, M.J. (1976) Geological mapping using Landsat data. *Photogrammetric Engineering and Remote Sensing* **42**, 325–337.
- Siegrist, A.W. (1980) Optimum spectral bands for rock discrimination. *Photogrammetric Engineering and Remote Sensing* **46**, 1207–1215.
- Süzen, M.L. & Toprak, V. (1998) Filtering of satellite images in geological lineament analyses: an application to a fault zone in Central Turkey. *International Journal of Remote Sensing* **19**, 1101–1114.
- Vane, G., Duval, J.E. & Wellman, J.B. (1993) Imaging spectroscopy of the Earth and other Solar System bodies. In: *Remote Geochemical Analysis: Elemental and Mineralogic Composition* (eds C.M. Pieters & P.A.J. Englert), pp. 121–143. Cambridge University Press, Cambridge.

- Podwysocki, M.H., Segal, D.B. & Abrams, M.J. (1983) Use of multispectral scanner images for assessment of hydrothermal alteration in the Marysvale, Utah, mining area. *Economic Geology* **78**, 675–687.
- Pratt, W.K. (1978) *Digital Image Processing*. Wiley, New York.
- Renz, Andrew, N. (ed.) (1983) Remote Sensing for the Earth Sciences. *Manual of Remote Sensing*, 3rd edn, Vol. 3. Wiley, New York.
- Rothery, D.A. (1984) Reflectances of ophiolite rocks in the Landsat MSS bands: relevance to lithological mapping by remote sensing. *Journal of the Geological Society of London* **141**, 933–939.
- Rothery, D.A. (1985) Interactive processing of satellite images for geological interpretation. *Geological Magazine* **122**, 57–63.
- Rothery, D.A. & Francis, P.W. (1985) A remote sensing study of a sector collapse volcano. *Proceedings of the 18th International Symposium on Remote Sensing of Environment, Paris*.
- Rowan, L.C. & Bowers, T.L. (1995) Analysis of linear features mapped in Landsat Thematic Mapper and side-looking airborne radar images of the Reno 1° × 2° Quadrangle, Nevada and California: implications for mineral resource studies. *Photogrammetric Engineering and Remote Sensing* **61**, 749–773.
- Rowan, L.C., Goetz, A.F.H. & Abott, E. (1987) Analysis of Shuttle multispectral infrared radiometer measurement of the Western Saudi Arabian shield. *Geophysics* **52**, 907–923.
- Rubin, T.D. (1993) Spectral mapping with imaging spectrometers. *Photogrammetric Engineering and Remote Sensing* **59**, 215–220.
- Ruiz-Armenta, J.R. & Prol-Ledesma, R.M. (1998) Techniques for enhancing the spectral response of hydrothermal alteration minerals in Thematic Mapper images of Central Mexico. *International Journal of Remote Sensing* **19**, 1981–2000.
- Schowengerdt, R.A. (1983) *Techniques for Image Processing and Classification in Remote Sensing*. Academic Press, New York.
- Siegal, B.S. & Abrams, M.J. (1976) Geological mapping using Landsat data. *Photogrammetric Engineering and Remote Sensing* **42**, 325–337.
- Siegrist, A.W. (1980) Optimum spectral bands for rock discrimination. *Photogrammetric Engineering and Remote Sensing* **46**, 1207–1215.
- Süzen, M.L. & Toprak, V. (1998) Filtering of satellite images in geological lineament analyses: an application to a fault zone in Central Turkey. *International Journal of Remote Sensing* **19**, 1101–1114.
- Vane, G., Duval, J.E. & Wellman, J.B. (1993) Imaging spectroscopy of the Earth and other Solar System bodies. In: *Remote Geochemical Analysis: Elemental and Mineralogic Composition* (eds C.M. Pieters & P.A.J. Englert), pp. 121–143. Cambridge University Press, Cambridge.

6 Thermal Images

Beyond a wavelength of about $4\text{ }\mu\text{m}$, energy from the Earth's surface is almost exclusively the result of radiant emission from natural materials. Stefan-Boltzmann's Law relates the total energy emission to the fourth power of a material's absolute temperature. Wien's Displacement Law describes the spread of emission among a range of wavelengths, rising to a maximum determined by absolute temperature (Section 1.2, Fig. 1.3). Both laws relate to the ideal behaviour of blackbodies—perfect absorbers and emitters of radiant energy. Natural materials are not blackbodies, and they only approximate these relationships. The Earth's average surface temperature of 300 K means that its energy emission peaks at $9.7\text{ }\mu\text{m}$. Gases in the atmosphere absorb energy in distinct wavebands, so that only a small proportion of the emitted spectrum is transmitted. Incidentally, this lies behind the so-called 'greenhouse' effect of gases such as water, carbon dioxide and methane, that slow down heat loss by radiation and so allow the mean surface temperature to be 33 K above that likely if the planet had no atmosphere. The main atmospheric window for thermal remote sensing is between 8 and $14\text{ }\mu\text{m}$ (Fig. 1.5).

Thermal infrared imaging uses devices similar to those suitable for the reflected region (Sections 3.3 and 3.4), based on cooled solid-state detectors (Fig. 3.9). Generally the detectors give a lower signal-to-noise ratio than those used for shorter wavelengths, resulting in coarser spatial and spectral resolution. The earliest thermal images were of the analogue type, signals from a line scanner modulating the brightness along a line on a cathode-ray screen to build up an image from successive lines on film moving across the screen. Digital thermal images from modern systems can be enhanced and analysed by the image-processing methods described in Chapter 5.

One advantage of thermal infrared images is that the Earth's surface is always far above absolute zero, and emits thermal radiation equally well at night as it does during the day. The possibility of night-time surveillance drove the invention of thermal imaging devices by military scientists. Fortuitously this adds to the possibilities for geological interpretation. The reason for this is that different materials lose heat at different rates during the night, and show up in quite different ways compared with their daytime appearance. The only limitations are the degree of cloud cover, which thermal radiation cannot penetrate, and the atmospheric window of reduced attenuation.

The thermal properties of surface materials are very different from those that control energy transfer in the

reflected region of the EM spectrum. As a result thermal images are often poorly correlated with those from the visible and near-infrared range. They provide an extra dimension for the discrimination of different rocks, soils and vegetation. They can be used in three main ways for investigating geological features. First, they may be interpreted in just the same way as photographs, given an understanding of the thermal properties of the surface. Second, images of the same area taken at different times of day give semiquantitative information about how different materials respond to heating and cooling. This is the basis of estimating variations in thermal inertia (Section 1.3.2). The third application depends on quantum mechanics and laboratory investigations, which show that troughs in the emission spectra of minerals, related to Si–O bond stretching and other phenomena, occur at different wavelengths depending on chemical composition and molecular structure (Section 1.3.2). It relies, therefore, on gathering thermal data in narrow wavebands to highlight these distinctive features.

6.1 What a thermal image shows

The variations in tone or DN in a thermal image are measures of the surface's radiant emission (Section 1.3.2). In the 8– $14\text{ }\mu\text{m}$ waveband natural materials are efficient absorbers of solar radiation and very little is reflected (Fig. 6.1). If the surface was a blackbody the image tone or DN would be an accurate measure of the true surface temperature. Radiant emission is controlled by a property of matter known as emissivity (Section 1.3.2). A blackbody has an emissivity of 1 whereas all natural materials have values less than 1. Most natural materials are selective emitters rather than true greybodies and their emissivity varies with wavelength (Section 1.3.2). The temperature derived from thermal infrared measurements by applying the Stefan-Boltzmann Law—the apparent blackbody temperature of the surface—is therefore always less than the true surface temperature.

Materials with high emissivity absorb and emit larger proportions of incident energy and heat energy transmitted into them than those with low emissivity. Whether the temperature of a surface increases or decreases depends on increase or decrease in the **heat content of the surface material**. The law of energy conservation determines the heat balance at the surface. Heat flows to and from the surface by many different means. Geothermal heat flows from below by conduction

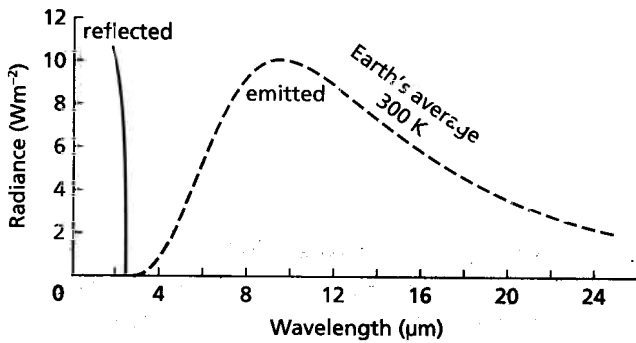


Fig. 6.1 The radiance of electromagnetic (EM) energy reflected by the Earth's surface falls rapidly towards zero beyond about $2.5 \mu\text{m}$. Absorption is virtually complete. Energy received from the Earth at wavelengths longer than this is almost entirely that emitted because of the temperature of the surface.

through rock and soil from the Earth's interior. Except in volcanic areas, this component is more or less constant over very large areas and far below the addition of solar energy. Solar radiation transmitted through the atmosphere is absorbed at all wavelengths, including a proportion in the visible and near-infrared region. Downward radiation contains a component of solar energy originally absorbed by the atmosphere at wavelengths between the atmospheric windows, but re-emitted at longer wavelengths. Energy is lost from the surface by an even more complex set of mechanisms. Most important is that radiating away from the surface, but significant losses involve the heating of air or evaporation of water involving latent heat at the ground surface and its movement by wind.

Radiative heat transfer dominates the surface heat balance. The principle of conservation of energy determines that for any radiation-matter interaction, the incident radiant flux at one wavelength (E_{λ}) is distributed between reflection (E_R), absorption (E_A) and transmission (E_T) by the material involved:

$$(E_{\lambda}) = (E_R)_{\lambda} + (E_A)_{\lambda} + (E_T)_{\lambda} \quad (1.5)$$

Dividing Equation (1.5) throughout by (E_{λ}) produces an expression allowing the spectral properties of the material to be defined in terms of the ratios (E_R/E_{λ}) , (E_A/E_{λ}) and (E_T/E_{λ}) , which are the spectral reflectance (ρ_{λ}), absorptance (α_{λ}) and transmittance (τ_{λ}), respectively, so that:

$$\rho_{\lambda} + \alpha_{\lambda} + \tau_{\lambda} = 1 \quad (1.7)$$

The vast bulk of geological materials are opaque and transmittance is zero, so that Equation (1.7) reduces to:

$$\rho_{\lambda} + \alpha_{\lambda} = 1 \quad (1.8)$$

Because good emitters are also good absorbers of radiation their emittances are equal to their absorptances,

and the two are interchangeable. Absorptance is difficult to measure and this relationship, known as Kirchhoff's Law after its initiator, can be transformed using Equation (1.8) to:

$$e_{\lambda} = 1 - \rho_{\lambda} \quad (1.9)$$

This theoretical relationship is amply illustrated by a motorist wearing shorts sitting on a black seat on a hot day. A white seat is distinctly more comfortable. The black surface has low reflectance and high absorptance—it heats up very efficiently—and vice versa. Most solar energy that penetrates the atmosphere is carried by visible and near-infrared radiation (Fig. 1.5a). Heating of the surface is therefore the result mainly of absorption of these wavelengths and their transformation to vibration of the material's component molecules. Solar radiation in the mid-infrared region is completely absorbed by any surface, but contributes relatively little to heating. A major factor in the heat balance of a surface therefore is its ability to reflect radiation over the whole visible and near-infrared range. This is the albedo of the surface, which controls its overall reflectivity.

To appreciate the surface heat balance and how the surface temperature of an area changes with time means considering the heat added to and removed from the surface. As geothermal heat is usually constant and low compared with the heating effect of solar radiation, it can be disregarded outside volcanic areas. The amount of solar radiation at all wavelengths that is available to heat the surface depends on:

- 1 the elevation of the Sun above the horizon—a function of latitude, time of year and time of day;
- 2 atmospheric conditions, such as cloudiness, humidity and air density—a function of weather and altitude;
- 3 the topography in relation to solar azimuth and elevation—slopes facing the Sun intercept more radiation than those facing away. If slopes are steep enough or the Sun is low enough there may be shadows that receive only radiation from the sky.

Three main processes control removal of heat from the Earth's surface.

- 1 Radiation of thermal infrared.
- 2 Heat transfer by convection—this depends on the movement of heated air, either by convection cells or wind. Cool air is heated by a hot surface, but equally a cold surface may receive heat from warm air. The amount of heat transferred by convective means depends on temperature gradients in the atmosphere, wind speed and turbulence, the area of surface exposed—this depends on the surface roughness—and on the density of the air and hence altitude.

3 Transfer of latent heat. Evaporation of water from the surface involves adding the latent heat of vaporization from the surface to the water to form water vapour. This depends on the same variables as convective heat

transfer, together with the moisture content of surface materials and the atmospheric humidity. It is complicated by the transpiration from any vegetation cover. The reverse is dew formation when latent heat is added to the surface as water condenses on it.

Clearly, interpretation of thermal images is a weighty matter. As well as the addition and loss of heat at the surface, the heat balance depends on the movement of heat just below the surface, how that changes the temperature of rocks and soils, and the time factor involved. In addition to the albedo and emissivity of the surface, there are other factors at work.

The ability of a volume of material to store heat is expressed by its thermal capacity (c). This is the amount of heat required to raise the temperature of a unit volume of material by 1°C . Thermal capacity is the product of specific heat and density (ρ) and is expressed in heat units per unit volume per degree Celsius. Specific heat varies very little for rocks and soils, so density differences have most effect on their ability to store heat. The higher the density the more heat can be absorbed by the surface with only a 1°C rise in temperature. Low-density materials undergo a greater change in temperature as heat is added or lost.

An expression of the ability of a material to pass heat by conduction is its thermal conductivity (K), defined as the amount of heat passing through unit area for a unit distance in unit time for a temperature difference of 1°C . Rocks and soils are poor conductors of heat, but vary over an order of magnitude. As water and air have even lower values of K , their presence in porous rocks extends the range further.

The rate of change in temperature in a material is controlled by its thermal diffusivity (κ). This combines thermal capacity, thermal conductivity and density in the relationship:

$$\kappa = K/c\rho \quad (6.1)$$

Thermal diffusivity is the most convenient measure of a material's ability to transfer heat from the surface to the interior during heating and vice versa.

Thermal inertia (P) is a measure of the resistance of a material to a change in its temperature in response to a change in the temperature of its surroundings, given by:

$$P = \sqrt{\rho c \kappa} \quad (6.2)$$

As the Earth's ambient temperature varies fairly regularly throughout the 24 h diurnal cycle, thermal inertia is the most appropriate measure of the combined thermal properties of surface materials. It is also relatively simple to estimate from digital thermal images captured at different times of day (Section 6.3.1). Values for thermal inertia together with other parameters measured for various surface materials in the laboratory are given in Table 6.1.

How the layer of soil and rock immediately beneath the surface heats and cools is a complex interplay between all these parameters. Heat flows from hotter to cooler regions in this layer at a rate determined by thermal conductivity. Daytime heating of the surface causes heat to be conducted downwards. Once surface cooling begins in the afternoon the heat stored at depth flows back to the surface. During the diurnal cycle the temperature profile in the top layer therefore changes. The shape of

| Material | Thermal conductivity, K ($\text{J m}^{-1}\text{s}^{-1} \text{K}^{-1}$) | Density, ρ (kg m^{-3}) | Thermal capacity, C ($\text{J kg}^{-1} \text{K}^{-1}$) ($\times 10^{-2}$) | Thermal diffusivity, κ ($\text{m}^2 \text{s}^{-1}$) ($\times 10^{-6}$) | Thermal inertia, P ($\text{J m}^{-2} \text{s}^{-1/2} \text{K}$) ($\times 10^{-3}$) |
|--------------|---|---|---|---|--|
| Peridotite | 4.6 | 3200 | 8.4 | 1.7 | 3.5 |
| Gabbro | 2.5 | 3000 | 7.1 | 1.2 | 2.3 |
| Basalt | 2.1 | 2800 | 8.4 | 0.9 | 2.2 |
| Granite | 3.1 | 2600 | 6.7 | 1.6 | 2.2 |
| Pumice (dry) | 0.3 | 1000 | 6.7 | 0.4 | 0.4 |
| Ignimbrite | 1.2 | 1800 | 8.4 | 0.8 | 1.3 |
| Serpentine | 2.6 | 2400 | 9.6 | 1.3 | 2.6 |
| Quartzite | 5.0 | 2700 | 7.1 | 2.6 | 3.1 |
| Marble | 2.3 | 2700 | 8.8 | 1.0 | 2.3 |
| Slate | 2.1 | 2800 | 7.1 | 1.1 | 2.1 |
| Shale | 1.8 | 2300 | 7.1 | 0.8 | 1.4 |
| Limestone | 2.0 | 2500 | 7.1 | 1.1 | 1.9 |
| Dolomite | 5.0 | 2600 | 7.5 | 2.6 | 3.1 |
| Sandstone | 5.0 | 2500 | 8.0 | 1.3 | 2.3 |
| Gravel | 2.5 | 2100 | 8.4 | 1.4 | 2.1 |
| Sandy soil | 0.6 | 1800 | 10.0 | 0.3 | 1.0 |
| Clay soil | 1.3 | 1700 | 14.7 | 0.5 | 1.8 |
| Water | 0.5 | 1000 | 42.3 | 0.1 | 1.5 |

Table 6.1 Thermal properties of rocks and water measured experimentally at 293 K.

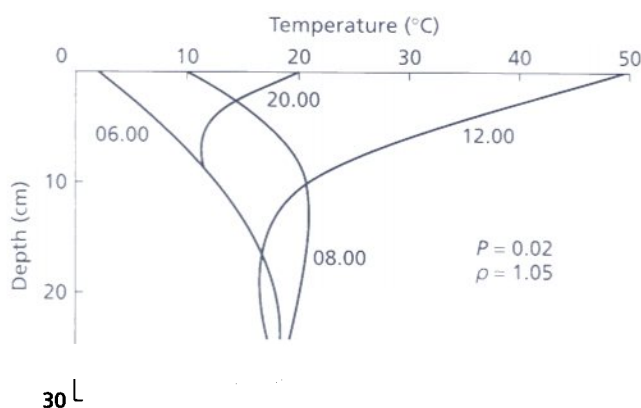


Fig. 6.2 The results of modelling the thermal behaviour of a thin layer at the Earth's surface show that temperature varies with depth. The profiles of temperature against depth change according to the time of day down to a depth of about 30 cm. The greatest variation occurs in the top 10 cm or so. Examining the curves in a time sequence shows that a heating 'wave' rises through the ground from dawn to noon. Cooling then progresses downwards, becoming slower during the night. The curves were modelled for relatively low thermal inertia and density, so that they represent extreme rates of heating and cooling. After Kahle (1980).

the profile and how it changes with time are governed by the surface heating and the various thermal properties of the materials involved.

Figure 6.2 shows a time sequence of temperature profiles modelled for varying surface heat input and a typical set of thermal parameters for a uniform, near-surface layer. The profiles show how temperature changes down the layer as a heating wave from dawn to the hottest part of the day, and as a wave of cooling as solar heating wanes after noon. Below about 30 cm, temperature stays the same throughout the daily cycle. For a given material, this depth varies over the course of a year because of seasonal variations in solar elevation and weather conditions. The shapes of such profiles vary with the values of the different thermal parameters of the near-surface materials. An increase in conductivity allows the heating wave to penetrate further, so distributing the solar energy through a greater volume. As a result, surface temperature does not rise so high or so quickly and changes more slowly with depth. An increase in thermal capacity—controlled mainly by density—means that more heat is required to raise the temperature of the topmost part of the layer. Less is left to penetrate to depth. Temperature at the surface does not rise so high or so quickly, but changes more rapidly with depth. Both these parameters are combined as a product in thermal inertia (Equation 6.2), and so conspire to magnify these two effects as thermal inertia increases. One result of this modelling is that although the shapes of profiles change for materials with different conductivities and thermal capacities, if they have the same thermal

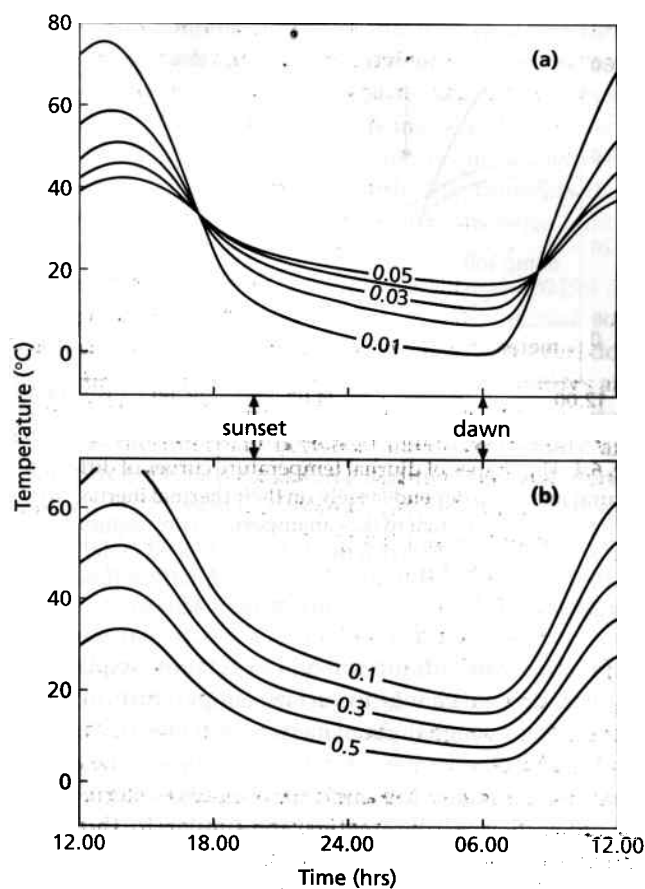


Fig. 6.3 The diurnal variations in temperature of surfaces underlain by materials of different thermal inertia (a) show different rates of heating and cooling. As a result, they cross one another about 2 h after dawn and 1 h before sunset. Temperature-time curves for similar materials with different albedos (b) are approximately parallel. Those with low albedo (dark surfaces) consistently have higher surface temperatures than those with higher albedos (light surfaces). Both sets of curves result from mathematical models rather than actual measurements. After Watson (1975).

inertia their surface temperatures at any time are the same. Because of this, thermal inertia can be estimated from digital images with a fair degree of confidence.

How surface temperatures vary through the diurnal heating and cooling cycle (Section 1.3.2, Fig. 1.14) depends on the thermal properties of the materials beneath. Figure 6.3 shows diurnal variations in surface temperature calculated for different values of two important thermal parameters—thermal inertia and albedo—with other parameters held constant. The family of curves for different thermal inertias show clear thermal cross-overs. The reason for them is that a material with a low resistance to change in temperature—low thermal inertia—heats up quickly to a high temperature during the day and cools in a similar fashion. Conversely, materials with high thermal inertia heat and cool more slowly to less extreme temperatures. Metals and any materials

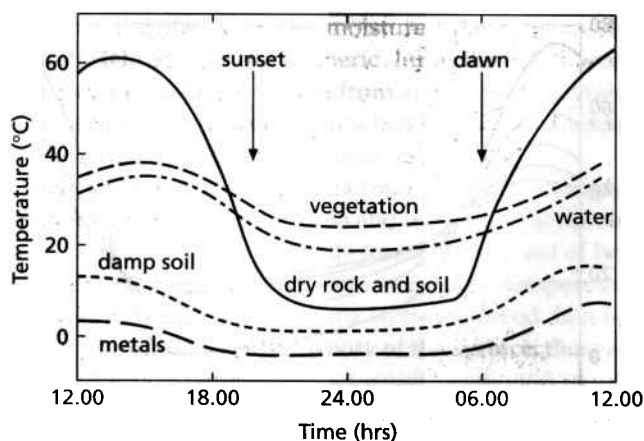


Fig. 6.4 The shapes of diurnal temperature curves of different natural materials depend largely on their thermal inertias, to which moisture content makes an important contribution. The actual temperatures that they achieve depend mainly on albedo and, to a lesser extent, on geothermal heat flow. The curves are merely sketches of relative behaviour.

containing water (damp soil and vegetation) experience the least diurnal change in surface temperature, because they have very high thermal inertias. Although the curves for different albedos have different positions on the graph, they are essentially the same shape. Curves derived for all other thermal parameters are similar to those for albedo. This confirms the importance of thermal inertia in mid-infrared remote sensing. Measuring the difference between day- and night-time surface temperatures gives a clue to thermal inertia. A large difference implies a low thermal inertia and a small variation suggests the converse. For materials with the same thermal inertia but very different values of all other thermal parameters this difference will not vary much.

Figure 6.4 is a sketch of diurnal temperature curves for a number of surface materials. It allows us to predict relative grey levels relating to these materials on thermal infrared images taken at different times of day (Section 6.2). It is also useful in selecting the times when greatest contrast exists between different materials.

6.2 Qualitative interpretation of thermal images

This section discusses day and night-time images of energy emitted in the 8–14 μm waveband, in either analogue or digital format, first as grey-tone images, and then combined with other data as false colour images.

6.2.1 Grey-tone thermal infrared images

Single-channel thermal infrared data produce grey-tone images. Like those of visible and near-infrared wavebands, the main elements of interpretation rely on tone,

texture, pattern, shape and context. Unlike images of shorter wavelength radiation, the tone of a thermal image expresses surface temperature and not reflectance. Cooler areas have dark tones and warmer areas appear light. The greatest and most common problem in interpreting images of reflected radiation occurs in areas of uniform albedo. They are typical of terrains with a thin skin of soil, vegetation or desert varnish. In such cases, thermal images are useful, because they should reveal differences in surface temperature related to various thermal properties to depths of 30 cm or so.

During daylight hours, as well as the differential heating effect of direct sunlight on materials with different thermal properties, images are affected by topography. Areas in shadow are cool and appear dark in much the same way as they do in more conventional images. Sunlit slopes are heated differentially according to their orientation. South-facing slopes are heated most strongly in the Northern Hemisphere. Exactly the same enhancement of topographic features normal to the Sun's azimuth, and suppression of parallel features occur as they do in images of shorter wavelengths. Daytime images express topography and geological structure very well, but in doing so obscure variations resulting from different thermal properties of the surface.

After sunset the effects of differential solar heating remain for a time. Loss of heat by radiation is controlled by both thermal inertia and surface temperature. For materials with the same thermal inertia, the hottest surfaces cool most quickly, and topographic effects are gradually removed. As Figs 6.3 and 6.4 show, surface temperatures at night become stable, particularly in the few hours before dawn. During the day temperature fluctuates rapidly as the Sun rises to its zenith and then begins to set. Because topographic effects are suppressed and temperatures are stable, night-time images are better for lithological interpretation. However, both types complement each other, when available. Figure 6.5 shows the different attributes of day and night-time images of the same area. As the properties that control the temperature of the surface are complex and unfamiliar, interpretation is best achieved with as much ground control as possible. The identity of some features, such as water bodies, can be suggested by their shape and image texture.

Planning the time of day to acquire images with the greatest information content relies very much on the nature of the problem. Figures 6.3 and 6.4 reveal that there are two brief periods during the day, 2–3 h before sunset and 2–3 h after dawn, when no matter what the variations in thermal inertia of materials surface temperature should be quite uniform, according to theory. These times of thermal cross-over are unsuitable for many applications. The same figures show that maximum temperature contrasts prevail during the hottest



Fig. 6.5 The top image shows the emitted infrared energy at 1100 hours from an area in the Front Range of Colorado, USA. The effect of solar shadowing is clear and highlights the local topography. The lower image is of the same area, but was acquired at 0400 hours, when surface temperatures are very stable. There is virtually no topographic effect and the image shows clear distinction between warm (bright) areas of dry outcrop and soil, and darker areas of cool damp soil. The bright patches on both images are small ponds. North is to the left in both cases and the scale is about 1 : 100 000. Courtesy of Ken Watson, US Geological Survey, Denver.

part of the day and before dawn. These times are those normally chosen for thermal imaging.

Dry rocks and soils have relatively low thermal inertias, whereas vegetation, standing water and wet soils have high values (Fig. 6.4). Merely to discriminate between these types of surface cover should be possible using day or night images. The effects of topography during daytime help to delineate geological structures, but hinder lithological discrimination. Although the contrast in surface temperatures over rocks of different thermal inertias is lower during the night, the suppression of topographic effects means that night images are best for discrimination.

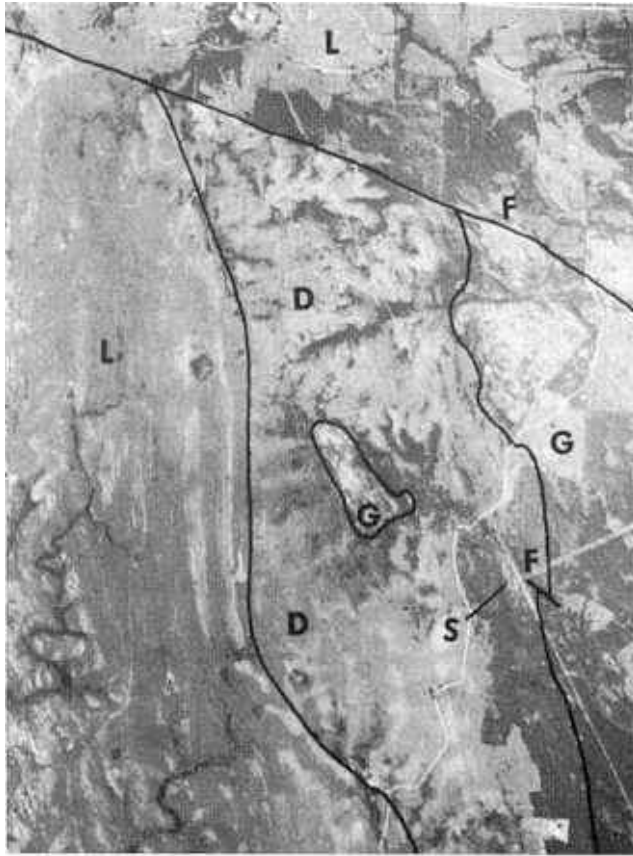
Variations in density affect the thermal inertia of dry rocks strongly (Table 6.1). A dense crystalline rock, such as peridotite, has a high thermal inertia. Its surface temperature remains fairly stable compared with less dense rocks, such as granites. The porosity of sediments and pyroclastic volcanic rocks mainly determines density and thermal inertia. A highly porous rock, such as pumice, displays rapid diurnal variations in temperature because of its low thermal inertia. Less extreme differences help distinguish between, for instance, sandstones with various degrees of cementation. The mineralogical composition of rocks also contributes to thermal inertia because of differences in thermal conductivity. The carbonates calcite and dolomite provide an interesting case. Dolomite is both denser and more conductive than

calcite, the combined effect being to give dolomites a thermal inertia roughly twice as high as those in limestones. Although the two carbonate rock types appear identical on shorter wavelength images, they are contrasted on thermal images (Fig. 6.6). As quartz has a very high conductivity compared with clay minerals there should be a good contrast between otherwise similar mudstones and siltstones.

These observations suggest that thermal images aid lithological discrimination where dry rocks and soils are near to the surface. Thermal inertia, however, does not entirely control the actual surface temperatures and image tones of different rocks. The other important factor is the interplay between albedo, emissivity and absorptivity. If two rocks have identical thermal inertia, but one has a higher albedo, areas underlain by that rock will show the lower temperatures and darker tones at any time of day. As rocks with other combinations of thermal inertia and albedo may well achieve similar temperatures, there is ample opportunity for confusion in only one image. The rocks in Fig. 6.6 all have similar albedo. Figure 6.7 shows examples of day and night thermal images from a dry terrain, where albedo varies markedly.

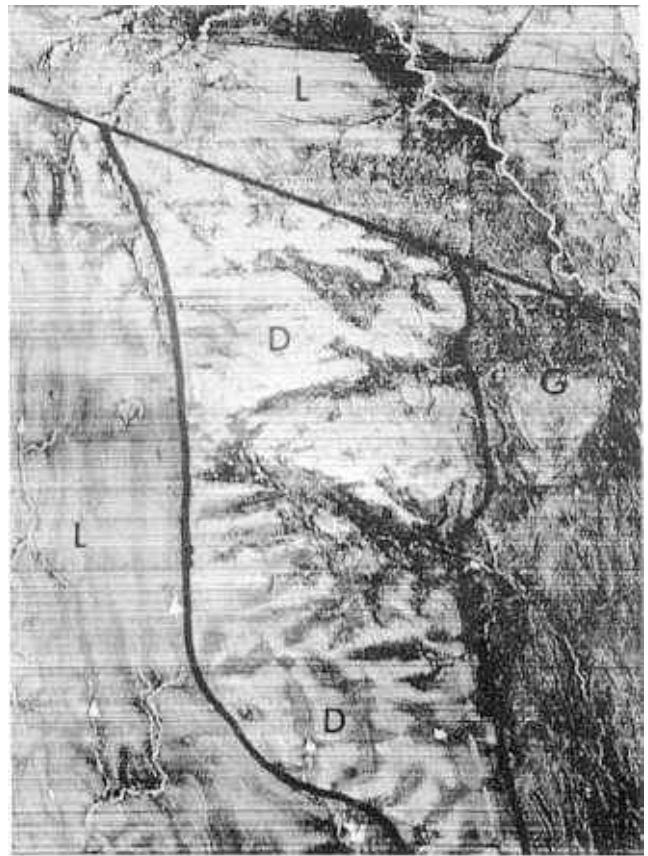
In humid terrains complications relating to water make their presence felt. The diurnal temperature curve for standing water on Fig. 6.4 does not have much to do with the thermal inertia of water, which in the laboratory is of the same order as those of rocks. The apparently high value in nature is caused by the fact that water is a liquid. It convects when temperature varies with depth. During the daytime the surface water in a lake achieves the highest temperature. This cools at night to be replaced by warmer water rising from below, thus stabilizing the surface temperature to some extent. This effect is magnified because water has a very high thermal capacity and therefore can store heat very efficiently. The net effect, as shown by Fig. 6.4 and illustrated by Fig. 6.6 (stream at top right), is that relative to dry rock and soil water appears cool (dark) during the day and warm (light) at night.

Vegetation introduces other complexities and anomalies. Because of its high water content green deciduous vegetation stores heat efficiently and appears warm on day images. During the night, water vapour is transpired from the leaves thereby shedding latent heat. This gives a relatively cool signature (Fig. 6.6). Overall the diurnal temperature pattern is similar to that of water (Fig. 6.4). Curiously, the composite effect of the needles on coniferous trees is to simulate a blackbody. This results in such vegetation having a brighter night-time signature than that during the day. Dead vegetation, such as grasses or agricultural stubble, has a low water content and a high content of air, both in stems and trapped between the foliage. It forms such a good insulating layer that



(a)

Fig. 6.6 The aerial photograph (a) of an area in Oklahoma, USA, shows mapped boundaries between limestone (L), dolomite (D) and granite (G). Apart from variations in texture, there is little to distinguish between the three rock types, and the dolomite appears identical to the granite. A thermal image acquired just before dawn (b) reveals contrasts between the three rock types with similar albedo, which are the result



of different thermal inertias. The high thermal inertia of the dolomite ensures that it remains warm during the night, whereas the limestone and granite, with lower thermal inertia, lose more energy and appear darker. Irregular dark patches in the dolomite represent wet soils in dry valleys. Courtesy of Ken Watson, US Geological Survey, Denver.

the ground retains heat and has a brighter night-time signature than bare soil.

Moisture contained in soils at or near the surface has a profound effect on the thermal behaviour of rocks. This has nothing to do with its intrinsic thermal properties but results from the cooling effect of evaporation. The more porous a rock is the greater the effect. Areas of damp terrain appear cool on images taken at any time (Fig. 6.6). The amount of cooling increases with air temperature and decreases as humidity rises. The greatest contrast between damp and dry soil therefore is during the daytime, particularly when the weather is hot and dry. Another enhancing factor is the speed of the wind, as this 'blow-dries' the surface, as well as increasing the convective loss of heat. Figures 6.8 and 6.9 illustrate some of the typical features of thermal images in humid terrains.

The thermal images shown up to this point are from airborne instruments that cover narrow swaths of ground at a large scale. Small-scale thermal images with

low resolution are also available from a variety of satellites. These include images from the Landsat TM with a resolution of 120 m. They are acquired at 0930 hours local time, which is close to the cross-over point on Fig. 6.3(a). As surface temperature shows the least contrast at that time and topographic effects are marked because of low Sun elevation, these images are of little use for geological interpretation—they appear like images of shaded relief. Although night-time TM thermal images can be acquired, this requires special requests, and in any case the overpass time is not predawn, but at around 2130–2230 hours, when differential heating effects have not had time to subside. Because land-surface and cloud temperature data are useful to meteorology, and global estimates of sea temperature are useful to fisheries, all meteorological satellites carry a thermal imaging system. That with the best resolution (1.1 km) is the AVHRR aboard the TIROS-N/NOAA series (Section 3.12.1). Since 1978 this has produced continuous day and night coverage of the whole globe at

4 km resolution and selective imaging at full resolution. The system has a ground swath about 2500 km wide and has great but underused potential for regional studies of geology and hydrology. An example of NOAA thermal imagery is shown in Fig. 6.10. One of the two constantly operating NOAA platforms has overpasses at 1430 and 0230 hours, which are better suited to geological interpretation of the thermal bands.

The Heat Capacity Mapping Mission (HCMM) (Section 3.12.5) was designed specifically for measurement of thermal inertia and thermal discrimination of surface materials. Although the amount of data and their coverage was limited by termination of the mission in 1980 after 2 years and data were received by only a few ground stations, HCMM data form a valuable resource for regional geological applications. The resolution is 600 m for swath widths of 716 km. Data were acquired during the day in a broad band between 0.55 and 1.1 μm to estimate surface albedo (Day-VIS), and in a 10.5–12.5 μm thermal band both at night (Night-IR) and in daytime (Day-IR). Because of the timing of day and night orbits, and variation in cloud cover, only a limited number of areas have both kinds of thermal image close enough in time to be compared meaningfully. Figure 6.11 shows examples of these. Although they reveal features of great geological interest, they are by no means as useful as combinations of the data with those from other systems, such as false colour composites (Section 6.2.2) and images of thermal inertia derived from the HCMM data (Section 6.3).

Fig. 6.7 (right) The area of Recent volcanic activity and intense desert erosion at Pisgah in the Mojave Desert, California, USA, is a classic test site for remote-sensing techniques. The digital image of reflected visible radiation in (a)—equivalent to albedo—shows very little information in the nearly black areas of lava flows, or in the very bright areas of superficial sediments and the evaporites in the dry lake at the bottom centre of the image. Images (b) and (c) are, respectively, of day- and night-emitted infrared. Comparison between the three images enables qualitative assessment of the effects of albedo on energy absorption, and of thermal inertia variations. For example, the high-albedo lake deposits are cooler during the day than the less reflective alluvial deposits surrounding them, as expected. The black lava flows have not achieved much higher daytime temperatures than the surrounding reflective sediments, however, which is contrary to first expectations. On the night image the lavas are much warmer than the sediments. This suggests that the lavas heat up and cool down more slowly than the sediments, and therefore have a higher thermal inertia. Some details of the lavas appear in the thermal images, but a fuller analysis is reserved for the images in Fig. 6.12 and Plate 6.3. The thermal images also reveal a clear linear feature running from top to bottom, just to the left of centre, which is not visible on the albedo image. This is probably an active fault that cuts both lavas and sediments. Courtesy of Anne B. Kahle, Jet Propulsion Laboratory, Pasadena.



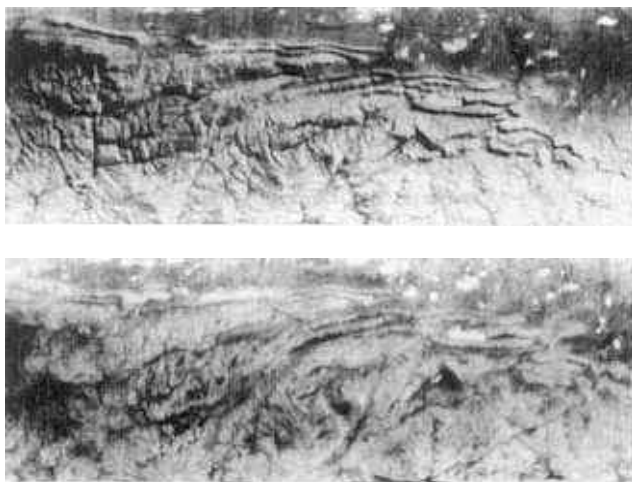


Fig. 6.8 The top of this pair of thermal images of the Front Range in Colorado, USA, was acquired at 1100 hours, the bottom just before dawn, and they show much the same contrasts as noted for Fig. 6.5. On both, but particularly on the day image, the complex folds at the mountain-plain boundary are clear, as are many linear features (probably faults) in the Rockies proper. The night image shows more clearly the different sedimentary units along the mountain front, and distinguishes between them and the dominantly granitic rocks of the Rockies in this area. The particularly bright folded layers at the front are quartzites, which have a very high thermal inertia. Courtesy of Ken Watson, US Geological Survey, Denver.

6.2.2 Using thermal data in false colour images

Despite the complexities of interpreting thermal infrared data in the form of grey-tone images, various methods have been devised to display them in colour. The attraction of doing this stems from the contrasted appearance of day and night images and the lack of correlation between thermal data and those from other parts of the spectrum. In theory, poorly correlated data displayed together as components of a colour image should produce a wide range of hues.

There are various options for colour display. The simplest is to use suitably contrast-stretched raw data to control red, green and blue components of an image. Despite the poor resolution the result in Plate 6.1 is a much stronger discrimination among the rock types present than in Fig. 6.11.

A means of adding fine topographic detail to low-resolution HCMM data is to combine them with better resolution data from another system. Merely to use, say, 80 m resolution Landsat MSS data to control one of the three primary colours adds little. The coarser HCMM data dominates the other two colour components, still resulting in an image with poor spatial resolution. To avoid this requires that the brightness or intensity of the image is controlled by the data of better spatial quality, while the colour incorporates thermal information. Plate 6.2 is an image using ISH colour space (Section 5.2.2). It



(a)



Fig. 6.9 This pair of daytime digital images contrasts reflected visible radiation (a) with emitted infrared (b) from an area of complex Precambrian and Lower Palaeozoic rocks in northern Scotland, which are largely blanketed by vegetation, boulder clay and peat. The albedo image shows very little detail. In contrast, the thermal image is

affected by subtle differences in vegetation, soil moisture content and thickness of the thin superficial veneer. This reveals several interesting structural features that would otherwise go unnoticed. Source: Open University, courtesy of NERC, Swindon.



Fig. 6.10 The TIROS-NOAA series of meteorological satellites provide coarse-resolution image data in the visible and near-infrared regions, and in the thermally emitted part of the spectrum, using the on-board AVHRR system. Image (a) shows AVHRR red reflectance (band 1) over part of south-eastern Iran, taken approximately 1 h after dawn. This is equivalent to an albedo image, and shows considerable variation over this tectonically complex area. The image of emitted radiation in the thermal band 4 (b) is much simpler. The most obvious feature is the zone of relatively cool rocks extending roughly north-south along the Dasht-e-Lut. This

represents an exotic block of old continental crust that was impacted into this geologically young terrain during the closure of Tethys. The younger, folded sediments of the Makran Range, running east-west in the south, and the less deformed sediments flanking the Lut Block have relatively high surface temperatures. Considering the time of day, the younger rocks possibly have higher thermal inertia than the crystalline rocks by the Lut Block, although there will be some confusion as a result of the time of the image being close to that of cross-over on Fig. 6.3(a).

combines Landsat MSS band 7 data with HCMM Day-IR and Night-IR.

A powerful means of colour display involving thermal data combines them with many sets of data from other parts of the spectrum. A convenient means of doing this is by principal component analysis (Section 5.4). In a display, those components loaded strongly by thermal data are combined with components with a strong loading for other data sets. Plate 6.3 is an image where red and green are controlled by PCs loaded with visible and near-infrared information, whereas blue is driven by a PC containing most of the thermal information. Although vivid in colour and clearly discriminating different kinds of surface, interpreting such images in terms of the spectral properties of different types of surface is difficult.

6.3 Semiquantitative analysis

Extracting information relating to surface thermal properties from images relies on their being in digital format. Semiquantitative analysis has two main objectives. Firstly, it exploits the diurnal changes in surface temperature at surfaces with different thermal properties to express them as thermal inertia. Secondly, it uses thermal information, either of one time or combined for day and

night as thermal inertia, together with data from other spectral regions to improve the separability and therefore classification of surface materials.

6.3.1 Thermal inertia estimates

Only sophisticated laboratory experiments can obtain accurate measurements of thermal inertia. Estimates using images derive from combining information on albedo and the temperature change of the surface during the diurnal heating cycle. Because it is impossible to measure the true surface temperature precisely using thermal data, the calculations use values of thermal inertia derived from the emitted energy, assuming that the surface acts as a blackbody. The mathematical relationship for this is fairly simple, and the conversion from DN to equivalent blackbody temperature is done by look-up tables for the instrument used. The average DN in the visible and near-infrared regions reflected from the surface during daytime provide albedo.

Although expensive, it is quite easy to acquire the three data sets needed for thermal inertia estimations using airborne line-scan systems. Finding suitable data from a satellite is by no means so simple. The only orbital system designed for this purpose was the HCMM. Its 600 m resolution data for albedo and day and night temperature are available only for areas covered by a



(a)

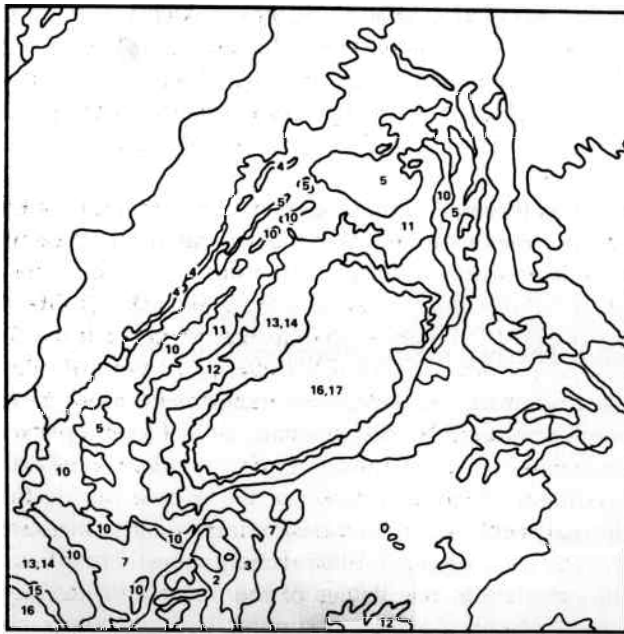
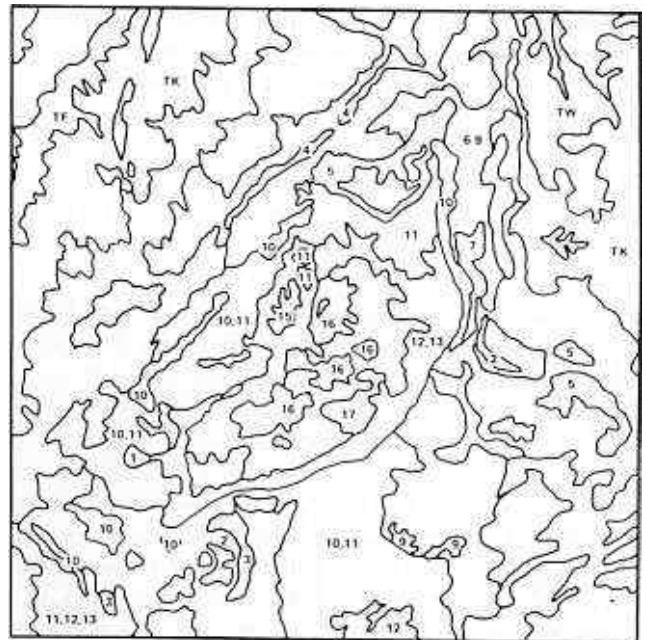


Fig. 6.11 Both images are of digitally processed HCMM data for the San Rafael Swell in Utah, USA, which is an eroded asymmetrical anticline. It exposes Permian sediments in its core, which are succeeded outwards by Triassic, Jurassic and Cretaceous sediments, mainly sandstones, shales and carbonates. Image (a) shows Day-IR data and (b) Night-IR data. Below each image is an interpretation based on the thermal data and the known stratigraphy. Units 17 and 16 comprise Permo-Triassic limestones and red shales. Unit 15 is a Triassic sandstone. Units 14, 13 and 12 are all pale, massive



(b)



sandstone of Triassic to Jurassic age. Units 11 and 10 are Jurassic shaley limestone and fine red sandstone. Units 5, 3 and 2 are Cretaceous shales, and unconsolidated Tertiary strata are represented by TF and TK. On each image all are crudely distinguishable. An important feature is that Units 12, 13 and 14 show cool day and warm night signatures, compared with the similar looking Unit 10, which is warm during the day and cool at night. Units 12, 13 and 14 have high thermal inertia, whereas Unit 10 has a low thermal inertia. Courtesy of Anne B. Kahle, Jet Propulsion Laboratory, Pasadena.

few ground-receiving stations. Global data are available at 1.1–4 km resolution from AVHRR. Although Landsat gathers thermal data, this is done routinely only during daytime overpasses. By special request it is possible to turn on the sensors at night. However, the unfortunate timings of the Sun-synchronous Landsat orbits (0930 and 2130 hours) coincide with the times of least contrast in surface temperatures (Fig. 6.3).

The first step in calculating the pixel-by-pixel variation of thermal inertia over an area is to register images of albedo, day and night thermal data. Accurate registration of the images from two times is difficult, particularly with airborne data that include relief displacement. Satellite data for day and night during the same 24-h cycle are captured from ascending and descending orbits, so that only lozenge-shaped areas where both ground tracks intersect are available for study. In the case of the short-lived HCMM experiment, the scope for study of thermal inertia is limited. The continuous, world-wide recording of thermal data from the NOAA AVHRR system offers much greater potential, although it has been little used.

The simplest measure of the rate of heat loss from the surface by radiation is the difference between day and night temperature (ΔT). This does not give an accurate estimate of thermal inertia because of the albedo effect. A surface with a high albedo heats up relatively slowly, has less heat to lose at night and so has a small value of ΔT . Low albedo surfaces soak up more heat, achieve higher day temperatures and lose heat very rapidly at night to give a high ΔT . A more meaningful estimate of apparent thermal inertia (ATI) combines albedo (a), ΔT and a correction factor for varying solar energy input to the surface (C), as follows:

$$ATI = C(1 - a)/\Delta T \quad (6.3)$$

The correction factor C depends on season, latitude, altitude and the angle and direction of slope. For most purposes corrections are made only for season and latitude. To remove topographic effects requires a digital image of elevation and extensive computations. In most cases the results show no striking improvement on uncorrected ATI. Values for a and ΔT are derived from the average visible to near-infrared reflectance, and from day and night temperature-estimates based on thermal image data. For HCMM these are Day-VIS, and Day-IR and Night-IR, respectively. The relevant AVHRR data are daytime Channels 1 and 2 for albedo, and Channels 4 and 5 for thermal data. The calculation is performed for each pixel in the registered scene of day and night data. The ATI values for each pixel are rescaled to the conventional 0–255 range, when ATI itself can be displayed as an image. Figure 6.12 compares images of ΔT and ATI derived from the aircraft data for the area shown in Fig. 6.7.

Like conventional day and night thermal images, from which they are derived, ATI images contain anomalous features. As well as being affected by topography, they contain features that result from variations in soil moisture content, varying wind speed, air temperature, humidity and various other factors. As it is not possible to correct for all these factors, interpretation must be backed up by surface observations on the day of the image acquisition.

6.3.2 Thermal data in classification

Images of thermally emitted infrared radiation and their derivatives express variations in density, porosity, moisture content and thermal inertia of surface materials. These have little if any effect on images of reflected radiation. Apart from the effects of narrow absorption bands, visible to near-infrared images are dominated by variations in albedo. Thermal infrared images contain an albedo effect, so that bright areas on an image of reflected energy are relatively dark (cool) on a thermal image. There is a negative correlation between the two types of image. The albedo effect is removed by the calculations involved in producing an ATI image and so they are more or less uncorrelated with images in the visible to near-infrared spectrum. Clearly, this improves the potential for discrimination between different types of surface.

Images of ATI allow discrimination in areas of uniform albedo, where visible and near-infrared images are least effective (Fig. 6.13). Examining the range of thermal inertias for some common rocks (Table 6.1 and Fig. 6.14) shows several interesting features that help orientate the use of thermal data in classification. Igneous rocks of widely different composition are not well separated by thermal data, with the exception of ultrabasic rocks. They are best discriminated by reflected radiation. Sedimentary rocks on the other hand show an excellent range of values, reflecting large variations in their porosity, density and mineralogy. Of particular importance is the strong contrast between limestone and dolomite, which are virtually indistinguishable by any other remote-sensing technique. Thermal data are also useful in distinguishing between crystalline, partly cemented and uncemented sediments, because of variations in porosity and density. Quartzite, sandstone and sand are composed mainly of quartz and therefore difficult to distinguish on images of short-wavelength radiation. They show strong contrast on thermal inertia images. Even where the data available do not permit calculation of thermal inertia values, as with Landsat TM band 7, a single thermal dimension can be used as a powerful discriminant in classification, simply because of low correlation with other dimensions.

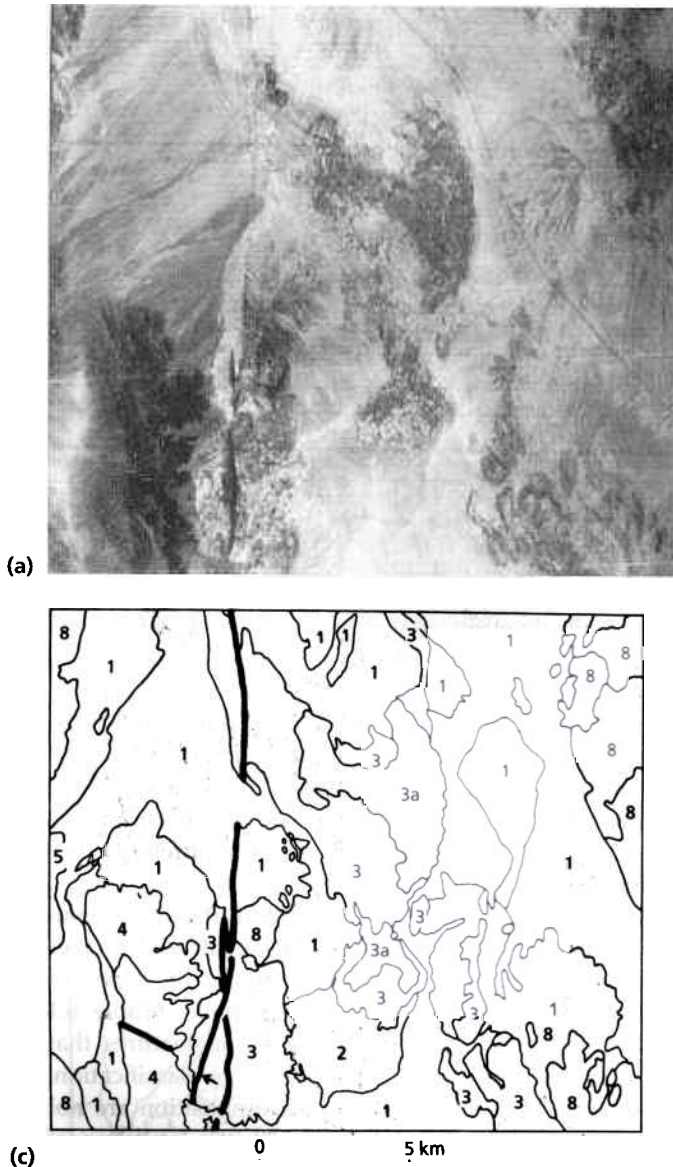


Fig. 6.12 Image (a) is of temperature difference between day and night derived from aircraft data of the same volcanic area in the Mojave Desert of California, USA, as Fig 6.7 and Plate 6.3. Image (b) is of apparent thermal inertia. The map (c) shows rough outlines of the major outcropping rock units. Unit 1 is alluvium, 2 represents saline flats on a playa lake, 3 is basalt (3a) with an aa surface), 4 indicates rhyolite intrusives, 5 is granite and 8 is andesite. The image helps to discriminate igneous rocks with high thermal inertia from alluvium with low, and identifies a division between aa and pahoehoe surfaces within the central Pisgah flow. However, it makes no clear distinction between igneous rocks such as 3, 4 and 8 with very different compositions. Courtesy of Anne B. Kahle, Jet Propulsion Laboratory, Pasadena.

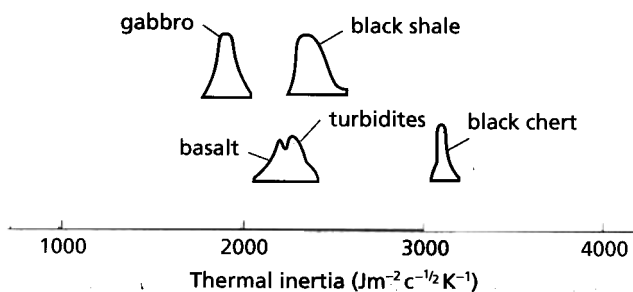


Fig. 6.13 Landsat MSS data fail to discriminate between rock types with similar, low albedos. In thermal imaging, constant albedo is a distinct advantage, as variations in surface temperature then relate only to differences in density, thermal capacity and conductivity. These differences are best expressed by thermal inertias, shown in these histograms, where the spread among dark rocks is very wide. Data from Watson (1978).

6.4 Multispectral thermal data

Within the range 8–14 μm the emission spectra of silicate minerals contain a prominent, broad absorption trough and associated features caused by Si–O bond stretching. The position of this trough and the shape of its shoulders are controlled by the co-ordination of silicon with oxygen. As the proportion of silicon in a silicate increases, the absorption feature shifts to shorter wavelengths. Because subdivision of silicates is according to Si–O co-ordination, this shift is potentially a means of discriminating between them using remote sensing (Section 1.3.2, Fig. 1.12). Similar features characterize carbonates (C–O bond stretching), iron minerals (Fe–O), clay minerals (Si–O–Si, Si–O, Al–OH) and various other groups of minerals

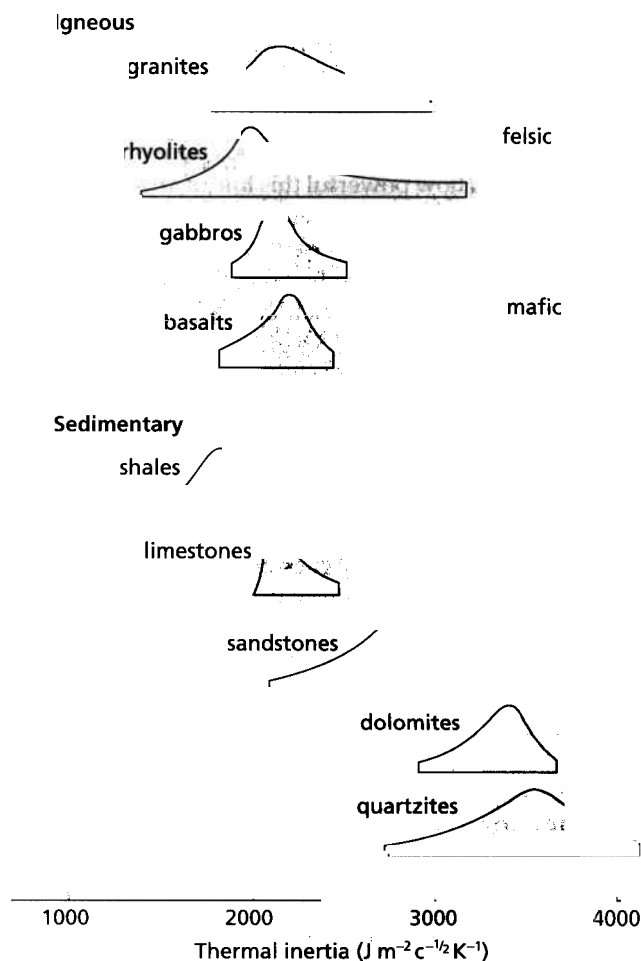


Fig. 6.14 A comparison between histograms showing the range of thermal inertia in igneous and sedimentary rocks shows that thermal imagery is more helpful in mapping sedimentary basins than areas of crystalline rock. Metamorphic rocks show much the same overlaps as igneous rocks. Data from Hummer-Miller and Watson (1977).

(Section 1.3.2, Fig. 1.13). The wealth of different, compositionally controlled features in the thermal infrared part of the spectrum suggests a bright future for multi-spectral analysis within the 8–14 μm band. It is the region most likely to provide direct discrimination between rock types, because the most prominent features relate directly to rock-forming minerals rather than minor components such as limonite and clay minerals. Figure 6.15 shows spectra for a wide range of igneous and sedimentary rocks.

Despite the great attraction of multispectral thermal data, the difficulties in separating narrow wavebands in the emitted region has enabled only two experimental imaging devices to be produced to date—the TIMS and the Geoscan systems. Their spectral ranges are shown in Table 3.2. Both have been operated only from aircraft, although the ASTER instrument incorporates five narrow bands (Table 3.4) in the mid-infrared. ASTER is aboard the Terra satellite launched in December 1999.

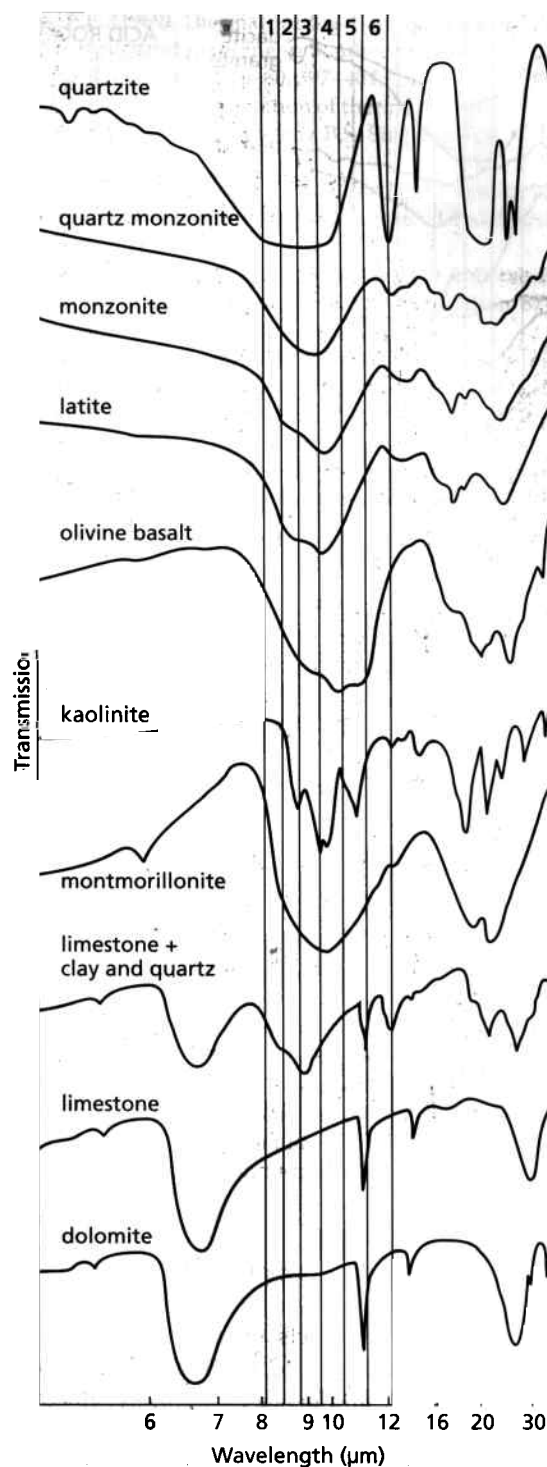


Fig. 6.15 These thermal-infrared transmission spectra of igneous and sedimentary rocks depend on the component minerals in each rock (Figs 1.12 and 1.13). The vertical bars indicate the wavebands detected by the experimental Thermal Infrared Multispectral Scanner (TIMS). The spectra are stacked to show relative shapes. From Kahle and Rowan (1980).

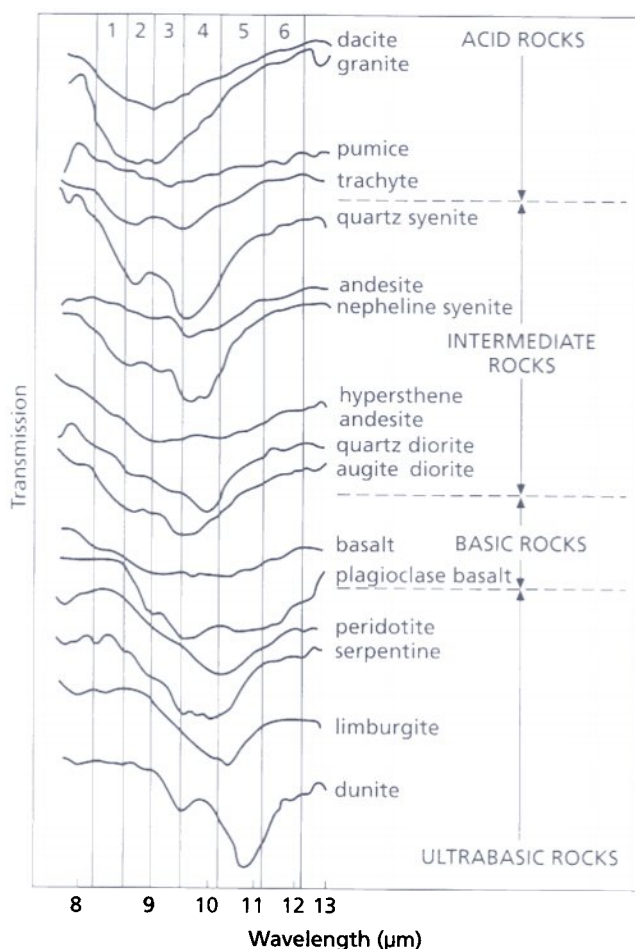


Fig. 6.16 Falling SiO_2 content of igneous rocks (and metamorphic rocks of the same range of compositions) results in a progressive shift of the Si-O bond-stretching absorption feature to longer wavelengths in thermal emission spectra. The positions of the TMS channels (vertical bars) suggest that multispectral thermal studies could become a powerful tool in geological mapping of crystalline terrains. The spectra are stacked to show their shapes more clearly. Data from Vickers and Lyon (1967).

In theory, choosing an appropriate combination of bands based on laboratory spectra and suitable ratios should highlight the effects of different minerals. In practice, the radiance emitted by the surface is dominated by temperature and hence topography, albedo and thermal inertia. There is a very high degree of correlation between thermal bands. Colour composites of raw bands or ratios, even with contrast stretching, are very bland. The best approach is using the decorrelation stretch discussed in Section 5.4. This fully exploits colour space and produces images in which colours can be related to laboratory spectra and interpreted in terms of different rock types.

Experimental work on a number of test sites has shown up several interesting results, a few of which are highlighted by Plates 6.4 and 6.5. There seem to be two

major roles for this approach. The first relates to igneous rocks, which of course contain a blend of various silicate minerals. Figure 6.16 shows the progressive shift of the Si-O bond-stretching feature in rock spectra towards longer wavelengths as bulk silica content decreases. Plate 6.4 shows how powerful this line of attack could be. If ultrabasic rocks were present in the area even better results would have been possible. The other approach concentrates upon the differences in quartz, clay mineral and carbonate spectra and their effects on the signatures of sedimentary rocks. Plate 6.5 gives an example of sedimentary discrimination using this method.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file *Menu.htm* in the *Imint* folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

On the Menu page, select the link *Additional Images: Image Data*, where you will find examples of images based on thermally emitted data.

Further reading

- Abrams, M.J., Kahle, A.B., Palluconi, F.D. & Schieldge, J.P. (1984) Geological mapping using thermal images. *Remote Sensing of Environment* **16**, 13–33.
- Byrne, G.F. & Davis, J.R. (1980) Thermal inertia, thermal admittance and the effect of layers. *Remote Sensing of Environment* **9**, 295–300.
- Eberhardt, J.E., Green, A.A., Haub, J.G., Lyon, R.J.P. & Prior, A.W. (1987) Mid-infrared remote sensing systems and their application to lithological mapping. *IEEE Transactions, Geoscience and Remote Sensing* **GE-25**, 230–236.
- Hummer-Miller, S. & Watson, K. (1977) Evaluation of algorithms for geological thermal inertia mapping. *Proceedings of the 11th International Symposium on Remote Sensing of Environment*, pp. 1147–1160.
- Kahle, A.B. (1980) Surface thermal properties. In: *Remote Sensing in Geology* (eds B.S. Siegal & A.R. Gillespie), pp. 257–273. Wiley, New York.
- Kahle, A.B. & Abbot, E., eds (1986) *The TMS Data Users Workshop*. Technical Publications, Jet Propulsion Laboratory, Pasadena, CA.
- Kahle, A.B. & Goetz, A.F.H. (1983) Mineralogic information from a new airborne thermal infrared multispectral scanner. *Science* **222**, 24–27.
- Kahle, A.B. & Rowan, L.C. (1980) Evaluation of multispectral middle infrared aircraft images for lithologic mapping in the East Tintic Mountains, Utah. *Geology* **8**, 234–239.
- Kahle, A.B., Abrams, M.J., Alley, R.E. & LeVine, C.J. (1981) *Geological Application of Thermal Inertia Imaging using HCMM Data*. Publication 81–55, Jet Propulsion Laboratory, Pasadena, CA.
- Kahle, A.B., Palluconi, F.D., LeVine, C.J., Abrams, M.J., Nash, D.B., Nash, R.E. & Scheildge, J.P. (1983) *Evaluation of Thermal*

- Data for Geological Applications*. Publication 83-56, Jet Propulsion Laboratory, Pasadena, CA.
- Nash, D.B. (1985) Detection of bedrock topography beneath a thin cover of alluvium using thermal remote sensing. *Photogrammetric Engineering and Remote Sensing* **51**, 77-88.
- Ninomiya, Y., Matsunaga, T., Yamaguchi, Y., Ogawa, K., Rokugawa, S., Uchida, K., Muraoka, H. & Kaku, M. (1997) A comparison of thermal infrared emissivity spectra measures *in situ*, in the laboratory, and derived from thermal infrared multispectral scanner (TIMS) data in Cuprite, Nevada, USA. *International Journal of Remote Sensing* **18**, 1571-1582.
- Offield, T.W. (1975) Thermal-infrared images as a basis for structure mapping, Front Range and adjacent plains in Colorado. *Geological Society of America Bulletin* **86**, 495-502.
- Oppenheimer, C. & Francis, P.W. (1997) Remote sensing of heat, lava and fumarole emissions from Erta' Ale volcano, Ethiopia. *International Journal of Remote Sensing* **18**, 1661-1692.
- Pohn, H.A., Offield, T.W. & Watson, K. (1974) Thermal inertia mapping from satellites—discrimination of geological units in Oman. *U.S. Geological Survey Journal of Research* **2**, 147-158.
- Pratt, D.A. & Ellyett, C.D. (1979) The thermal inertia approach to mapping of soil moisture and geology. *Remote Sensing of Environment* **8**, 151-168.
- Rowan, L.C., Offield, T.W., Watson, K., Cannon, P.J. & Watson, R.D. (1970) Thermal infrared investigations, Arbuckle Mountains, Oklahoma. *Geological Society of America Bulletin* **81**, 3549-3562.
- Sabins, F.F. (1969) Thermal infrared imagery and its application to structural mapping in Southern California. *Geological Society of America Bulletin* **80**, 397-404.
- Sabins, F.F. (1980) Interpretation of thermal infrared images. In *Remote Sensing in Geology* (eds B.S. Siegal & A.R. Gillespie), pp. 275-295 Wiley, New York.
- Short, N.M. & Stuart, L.M. (1982) *The Heat Capacity Mapping Mission (HCMM) anthology*. NASA Special Publication 465. Washington DC.
- Vane, G. (ed.) (1987) *Proceedings of the 3rd Airborne Imaging Spectrometer Data Analysis Workshop*. Publication 87-30, Jet Propulsion Laboratory, Pasadena, CA.
- Vickers, R.S. & Lyon, R.J.P. (1967) Infrared sensing from spacecraft—a geological interpretation. *Proceedings, Thermophysics Special Conference*, American Institute of Aeronautics and Astronautics, Paper 67-284.
- Watson, K. (1975) Geological applications of thermal infrared images. *IEEE Transactions, Geoscience and Remote Sensing* **GE-63**, 128-137.
- Watson, K. (1978) Thermal phenomena and energy exchange in the environment. In: *Mathematical and Physical Principles of Remote Sensing*, pp. 109-174. CNES, Toulouse.
- Wolfe, E.W. (1971) Thermal IR for geology. *Photogrammetric Engineering and Remote Sensing* **37**, 43-52.
- Wood, J.A., Lasserre, M. & Fedosejevs, G. (1990) Analysis of mid-infrared spectral characteristics of rock outcrops and an evaluation of the Kahle model in radiative transfer.

7 Radar Remote Sensing

The microwave region of the EM spectrum presents two opportunities for gathering remotely sensed data. First, like radiation in the 8–14 μm range, the Earth's surface emits microwaves as a result of its temperature, according to the Stefan–Boltzmann relationship and Wien's Law (Section 1.2). Second, microwave radiation can be generated artificially as coherent waves. Not only does the atmosphere fail to attenuate microwaves to any great extent beyond about 0.3 cm, but electronic systems can detect very low energies. The first opportunity relates to passive microwave remote sensing, the second makes possible radar imaging, the only operational active remote-sensing system. Radar plays the dominant role in this chapter because passive microwave methods, although theoretically promising, have failed to produce useful results for geologists.

Radar imaging possesses many advantages relating to the control that is possible with an active system. The principles were covered in detail in Section 3.5. It too has a day or night, all-weather capability and can be directed at the surface or at the atmosphere. It depends on the scattering of energy by materials that lie in the path of an artificial microwave beam, known as back-scatter, and reception of the proportion of energy that returns back to the antenna. Radar waves are generated in coherent form, that is as radiation with a single wavelength where all waves are in phase. Coherent radar makes possible the use of Doppler shift in generating synthetic-aperture radar (SAR) images, where resolution is fine and independent of range (Section 3.5). Any microwave wavelength can be generated with any polarization. The angle of illumination of the surface can be varied by adjusting the antenna. Simply changing the flight line of the platform changes the direction of radar propagation. These controls help optimize illumination conditions for many kinds of application. Flexibility and complete control is essential, because the interactions between radar waves and the surface are both complex and very different from those for shorter wavelengths.

7.1 Interactions between radar and surface materials

What happens to the electromagnetic energy in a radar pulse when it meets the surface depends on four main factors:

1 the attitude of the surface—dealt with at length in Section 3.5;

2 the roughness and heterogeneity of the surface and of subsurface materials;

3 the wavelength, polarization and depression angle of the radar, which are controlled variables;

4 the electrical properties of the surface—the complex dielectric constant of surface materials.

In decreasing order of importance, all help determine the proportion of incident microwave energy that the surface scatters back directly to the antenna aboard an aircraft or orbiting platform. This governs the tone of the radar image. The brighter the tone the larger the back-scattered energy. As with images acquired from any part of the spectrum, the fundamental basis of interpretation lies in the spatial distribution of tones to give familiar textures, shapes and patterns. The tone at any point gives information relevant to the actual nature of the surface itself. This section considers the effects on back-scatter of variations in surface electrical properties and roughness and how these effects change with wavelength, polarization and depression angle. Going further means introducing a couple of important definitions.

A measure of the intensity of back-scattered energy from a point target is its radar cross-section. This is the area of a hypothetical surface that scatters radar energy equally in all directions and which would return the same energy to the antenna as the real point target. A measure of the energy back-scattered from a target with a large area, such as a field, is its radar scattering coefficient. This is the average radar cross-section per unit area. It is a dimensionless quantity, and as it varies over several orders of magnitude we express it as 10 times its logarithm, in decibels (dB). The radar scattering coefficient is a fundamental measure of the radar properties of a surface, and governs the tone of the surface on a radar image.

7.1.1 Complex dielectric constant

A material's dielectric constant controls the proportion of radar energy reflected by a material and that penetrating into it. Materials with a high dielectric constant, such as metals and water, are excellent reflectors and absorb very little energy. The lower the dielectric constant the more energy is absorbed and can re-emerge, giving the potential for penetration beneath the surface. The dielectric constant is not a simple function of wavelength, but varies relatively little for most rocks and soils over the range of commonly used wavelengths.

Dry soils and rocks have dielectric constants in the range 3–8. These are low enough to permit penetration

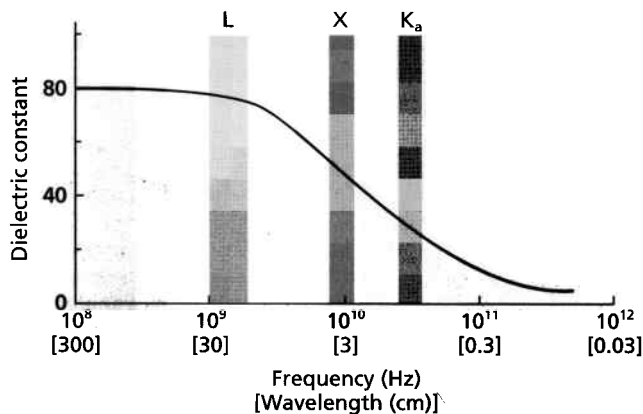


Fig. 7.1 As the wavelength of microwave radiation increases, so too does the effective dielectric constant of water. At the wavelengths of Ka- and X-band radar the dielectric constant is between 30 and 45, whereas for L-band radar it is at its maximum of about 80. Rocks and soils rarely contain more than about 15% water, so the effect of moisture on their dielectric properties is considerably reduced compared with that suggested by the graph. It might appear from the graph that, because the dielectric constant of water decreases at shorter wavelengths, Ka-band radar is most likely to penetrate beneath the surface. However, the physics of the phenomenon show that roughly an equal number of radar waves penetrate material with the same moisture content, irrespective of wavelength. So, contrary to expectations, it is L-band radar that is most effective at revealing subsurface features. Experiments indicate that unless the moisture content is less than 1%, penetration is negligible.

of a significant proportion of incident radar energy into dry materials to depths up to 6 m using L-band or 25 m radar. Penetration is only possible, however, where the topographic surface is radar-smooth. Subsurface returns emerge only when the subsurface has a radar-rough component (Section 7.1.2). As water has a maximum dielectric constant of 80, the greatest cause of variation in this parameter in rocks and soils is their moisture content. As moisture content increases the dielectric constant increases in a roughly linear fashion. However, the dielectric constant of water increases as wavelength increases (Fig. 7.1), and this has an effect on the radar-related properties of damp soil too. As even a saturated soil contains relatively little water the effect is not marked.

Penetration of radar into soil is expressed in terms of the number of wavelengths. The complex equations governing this behaviour suggest that for constant dielectric properties the same number of waves penetrate, irrespective of wavelength. Thus the longer the wavelength the greater the penetration. Recent studies indicate that moisture content greater than 1% effectively rules out any penetration, so this phenomenon can be exploited only in hyperarid terrains of terrestrial deserts and on planets such as Venus and Mars.

Because the leaves of growing plants have a high moisture content, vegetation has a higher dielectric constant than dry natural materials. Plants therefore are excellent radar reflectors. As a result only a proportion of radar energy is able to penetrate to the surface beneath, depending on the nature and density of the vegetation canopy. Figure 7.1 suggests that longer wavelength radar is more capable of penetrating vegetation to give information about the surface than that of short wavelength.

Except for its effect on penetration, dielectric constant is a minor factor in controlling the tone and texture of radar images. They are dominated primarily by slope effects (Section 3.5) and by surface roughness.

7.1.2 Roughness

A perfectly smooth surface of a material with a high dielectric constant acts as a mirror to radar, as it would to all forms of radiation. Being directed to the side of the platform, radar pulses meet a horizontal surface at an acute angle and are reflected away from the antenna at the same angle, without being scattered. This specular reflection results in a totally black signature for a smooth surface (Fig. 7.2a). Where a smooth surface is at right angles to the radar beam, the reflection is directly towards the antenna, giving an intensely bright response. Sand dunes and ocean waves commonly show this effect. Smooth reflectors with facets arranged at right angles result in multiple reflections between the facets. This is the principle behind the corner reflectors placed at the mastheads of small boats, so that they are detectable by navigational radar. Regardless of the angle at which a radar wave enters the cavity of a corner reflector, reflections between the facets ensure that the energy is returned directly to the antenna (Fig. 7.2b). Corner reflectors occur naturally at the intersections between bedding and rectangular jointing in rocks such as sandstones, limestones and lavas. They give rise to bright speckles, or blooms, within the area of more uniform tone that characterizes such rocks.

A rough surface is made up of countless irregularities, some of which mimic corner reflectors, whereas others give rise to more complex interactions. The net effect is to scatter radar energy diffusely in all directions. Some returns to the antenna as a measurable signal (Fig. 7.2c). Roughness, for the purpose of interaction with EM radiation is a relative term depending on wavelength and angle of incidence. Two relationships (Equations 7.1 and 7.2) allow roughness to be quantified. In them the height of irregularities—the roughness (h)—is determined by the wavelength λ and the angle of incidence θ . A surface that appears smooth to radiation satisfies the Rayleigh criterion:

$$h < \lambda / 25 \sin \theta. \quad (7.1)$$

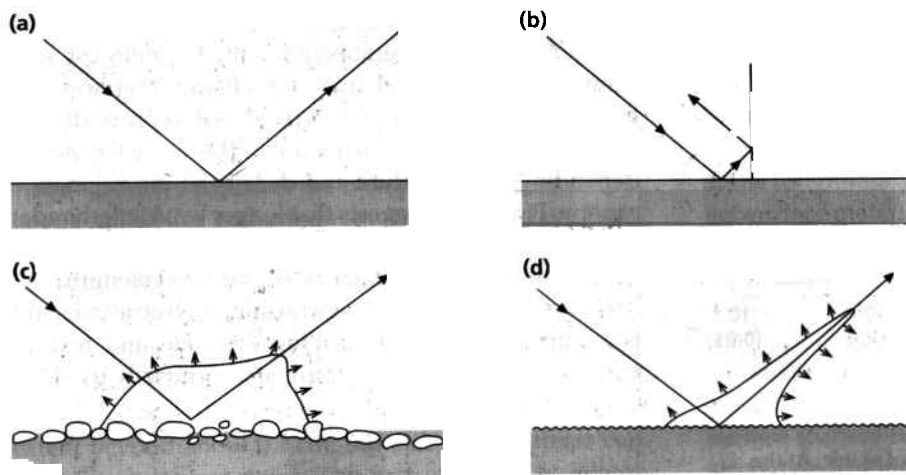


Fig. 7.2 How radar is back-scattered depends on the incidence angle and on surface roughness. Four possibilities are shown: (a) for a perfectly smooth horizontal surface; (b) for a corner reflector; (c) for a rough surface; and (d) for a smooth natural surface.

Table 7.1 Surface roughness (in cm) relative to radar wavelengths, derived from the Rayleigh criterion.

| Roughness | Ka-band ($\lambda = 0.86$ cm) | X-band ($\lambda = 3$ cm) | L-band ($\lambda = 25$ cm) |
|--------------|-----------------------------------|-------------------------------|--------------------------------|
| Smooth | <0.05 | <0.17 | <1.41 |
| Intermediate | 0.05–0.28 | 0.17–0.96 | 1.41–8.04 |
| Rough | >0.28 | >0.96 | >8.04 |

One that is rough satisfies the criterion:

$$H > \lambda / 4.4 \sin \theta \quad (7.2)$$

The behaviour of radar waves at natural surfaces is complex. Rough surfaces scatter energy diffusely in all directions, whereas surfaces of intermediate roughness combine specular and scattered components (Fig. 7.2d). The tone of a horizontal natural surface on a radar image therefore is a combined result of its roughness and to a much lesser extent of the dielectric constant of the materials from which it is formed.

Quite clearly the Rayleigh criterion implies that radar wavelength helps determine what is rough or smooth. Table 7.1 shows limiting values of the mean height of surface irregularities associated with different categories of roughness for three radar wavelengths in common use. They were derived using the Rayleigh criterion. Roughness also depends on the angle of incidence of radar waves at the surface. For radar images gathered from aircraft this is a major problem. Because the depression angle changes markedly from near to far range (Section 3.5), ensuring that the antenna can record from a wide swath of surface, means that the incidence angle changes too. The apparent roughness of a surface of uniform roughness therefore changes across the swath. As only a narrow range of depression angles can encompass a very broad swath from orbital altitudes, this effect on image tone is not apparent on Seasat and SIR images (Section 3.12.1). Depressor angle can be varied systematically on orbital SAR systems

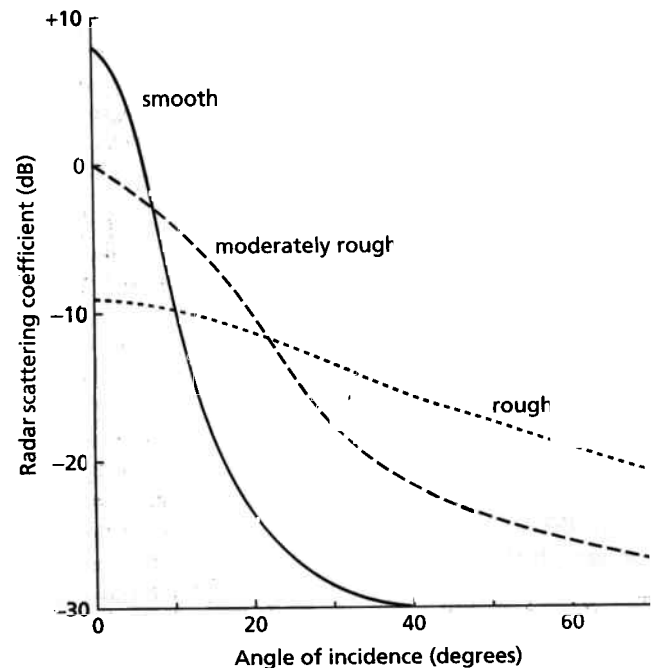


Fig. 7.3 The proportion of emitted energy back-scattered from a surface to the antenna—the radar scattering coefficient—depends on both the surface roughness and the angle of incidence. Both the depression angle and the surface slope control the incidence angle.

It is possible theoretically to exploit the variation of apparent roughness with incidence angle to gain more detailed information about the nature of the surface, or to select a depression angle that gives the greatest discrimination between surfaces of different roughness.

Figure 7.3 shows how the radar scattering coefficient depends on surface roughness and incidence angle. For low incidence angles smooth surfaces reflect a large proportion of energy directly back to the antenna. As the incidence angle increases, more and more energy is reflected away from the antenna, until at about 40° al

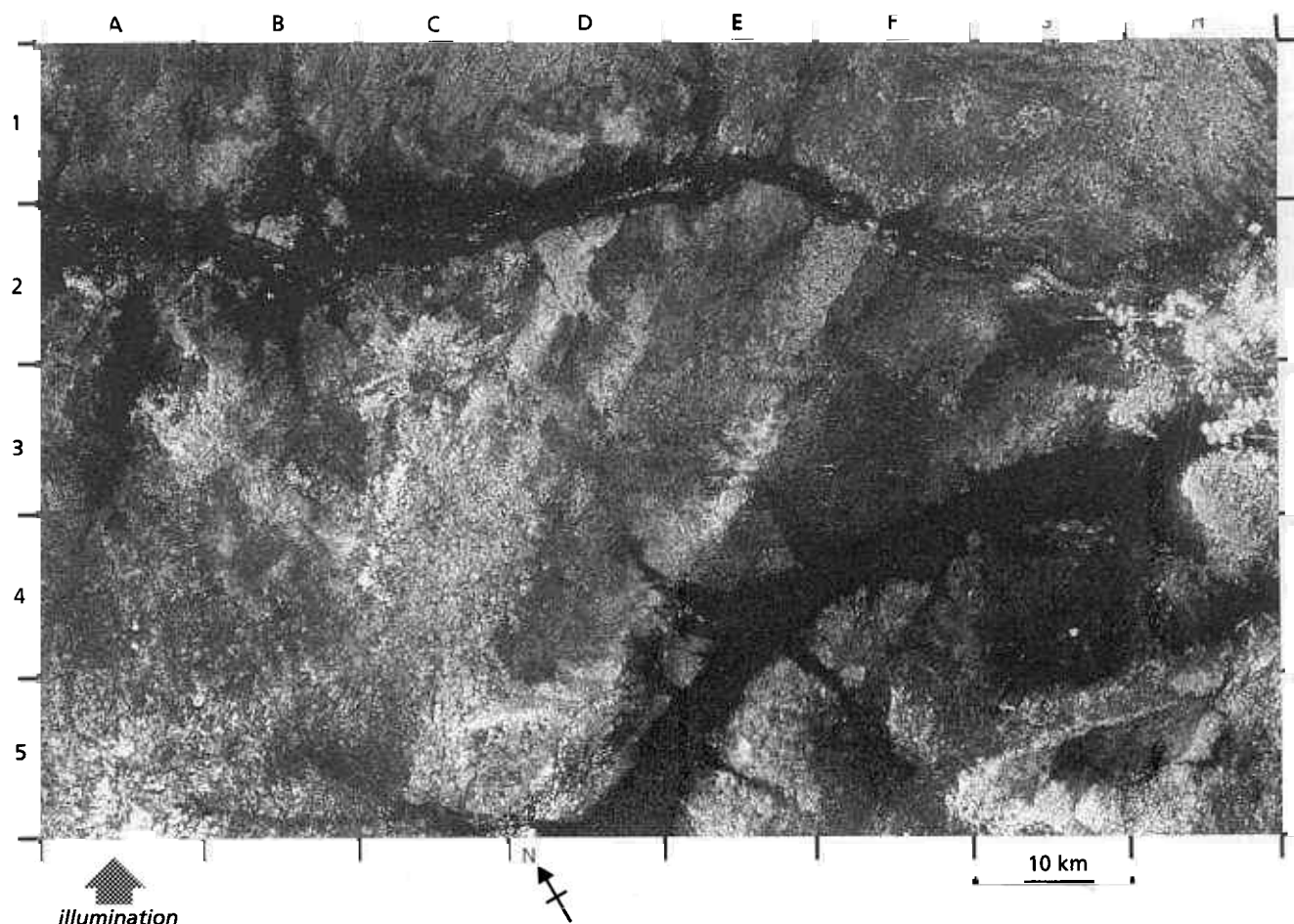


Fig. 7.4 This SIR-C image of part of the Sahara Desert shows the ability of radar to penetrate dry sediments. Landsat images show that the area is a nearly featureless sand and gravel plain. The image shows patterns which suggest a drainage network. Because tributaries of the main channel 'V' towards the south, water that produced the network was flowing southwards. The Nile system, some 250 km to the east of this area, flows to the north, as do most of the widely spaced active wadis in this area. The drainage pattern is thought to date from a Holocene wet period and to have been buried beneath sand as climate

became hyperarid. Several nearby archaeological sites confirm that the area was humid enough 6000 years ago to support agriculture. One theory holds that the patterns result from radar penetration and back-scatter from subsurface variations in grain size. An alternative is that the radar is picking up subtle variations in surface roughness related to the old stream courses. This image from the Shuttle Imaging Radar programme, together with most of those illustrating this chapter, was produced by the Jet Propulsion Laboratory, Pasadena.

energy is reflected away. Smooth horizontal surfaces therefore show as bright on images taken with a steep depression angle and dark on those taken looking further to the side of the platform. As rough surfaces scatter energy diffusely in all directions, the proportion returning to the antenna is virtually independent of incidence angle. Their brightness on an image usually does not change much with the angle at which radar waves are received at the surface. They appear dark relative to smooth and intermediate surfaces when incidence angle is steep—depression angle shallow—but form bright features at all incidence angles less than about 20° . Figure 7.3 implies that images taken with a judicious selection of depression angles can give a powerful discrimination of variations in surface roughness of horizontal surface. This is very useful where the tonal variation of a single

image is controlled by an unknown variation of both roughness and dielectric constant.

As well as scattered energy from the ground surface or the top of a vegetation canopy, the radar returns from the Earth's surface contain components from within the soil and the vegetation canopy itself. These comprise the results of complex volume scattering, mainly multiple reflections from electrical heterogeneities within the soil and among plant structures such as branches. Soil layers sometimes reflect and scatter some of the energy penetrating the soil surface, as can the buried soil-rock interface. In many cases it is impossible to separate all the different components and all contribute to the overall roughness of the surface but, in relatively simple terrains, it is possible to identify effects that result from subsurface features (Fig. 7.4).

7.1.3 Polarization

Like visible light, radar transmission and reception can be in different modes of polarization. Polarized radiation has its electrical component of wave vibration in a single plane perpendicular to the direction of propagation. For radar transmission and reception this usually is arranged to be either horizontal (H) or vertical (V). The most common combination is to transmit horizontally polarized radar and receive the horizontally polarized component of back-scattered radiation. This is assigned the coding HH, and gives like-polarized images. Horizontal transmission and vertical reception (HV) produces a cross-polarized image. There are four possibilities—HH, VV, HV and VH. The most usual strategy is to transmit horizontally and receive both horizontal and vertical components—HH and HV.

Interaction between the surface and radar waves usually leaves the sense of polarization unchanged. However, it can cause depolarization or rotation of the plane of polarization. So the mode of polarization can influence how objects appear on images. A proportion of the depolarized component of back-scatter and that rotated through 90° is detected in a cross-polarized image. How these effects are produced is not perfectly understood, but a widely held view is that some of these phenomena result from multiple reflections at or beneath the surface. Multiple reflections are at a minimum with bare rock or soil with a high dielectric constant. Where vegetation is thick and varied in structure there is a greater chance of multiple reflections between leaves, twigs and branches, and depolarized returns are common from vegetated surfaces. It is possible that subsurface back-scatter by soil particles or rock heterogeneities may give rise to similar effects.

7.2 Interpretation of radar images

Processes that produce the tonal and textural features on radar images are varied and complex. A general account of how the full range of rock types and geological structures respond to radar illumination requires a book in its own right. A mere impression of the possibilities is given in this section. The strategy adopted is first to apply some of the photogeological principles introduced in Chapter 4 to a series of radar images showing important geological phenomena (Section 7.2.1). Section 7.2.2 then discusses the advantages for structural geology of images taken with different radar look directions. Images taken with different depression angles reveal more information about surface roughness (Section 7.2.3). Pairs of radar images with different look directions or different depression angles introduce parallax, and this, as in overlapping aerial photographs, allows stereoscopic

viewing (Section 7.2.4). Multipolarized (Section 7.2.5) and multifrequency radar images (Section 7.2.6) provide additional information about the composition and texture of surface materials. Radar image data are increasingly available in digital form. They contain quantitative information and can be analysed by image processing techniques to extract concealed information (Section 7.2.7) or combined for comparative purposes with other kinds of data (Section 7.2.8).

7.2.1 Geological features on radar images

A radar image combines the attributes of tone, texture, pattern, shape, context and scale in much the same way as grey-tone images of shorter wavelength radiation. They differ in the way these interpretable attributes are produced, but the same photogeological principles apply (Chapter 4).

The most striking feature of a radar image is how it accentuates surface topography. To a large extent the amount of energy back-scattered to the antenna is governed by the attitude of slopes. Those perpendicular to the direction of wave propagation return nearly all incident radiation and appear bright. Those sloping away at a steeper angle than the radar are in shadow and show as totally black. Other slope angles produce returns in which intensity combines influences of slope together with those stemming from the properties of surface materials. Slope attitude produces layover and foreshortening (Section 3.5) so that topography is sometimes displayed in an unusual form. Despite this, radar's overall effect is to sharpen topography to a greater degree than seen in other kinds of image. In this respect, radar images are analogous to edge-enhanced, low Sun-angle images. Strong shadowing helps the eye appreciate topography because of the pseudostereoscopic effect (Section 2.5).

Topographic emphasis is most important for interpretation of geological structures underlying major destructional landforms. Their control is often the attitude of compositional layering and faults, folding, and the boundaries of igneous intrusions. Figures 7.5 and 7.6 show examples of this enhancement of layering and folds. Spectacular examples of topographic enhancement relate to faulting (Figs 7.7 and 7.8). Recognition of igneous intrusions often relies on tonal and textural differences as well as on the expression of cross-cutting contacts. Figure 7.9 gives some particularly good examples of both plutonic intrusions and dykes.

Another special property of radar aids enhancement of structures. Features with rectangular boundaries are natural corner reflectors (Section 7.1.2). All the radar energy that illuminates the angles returns to the antenna. Even where such a corner reflector is much smaller than a resolution cell, its back-scatter can saturate the

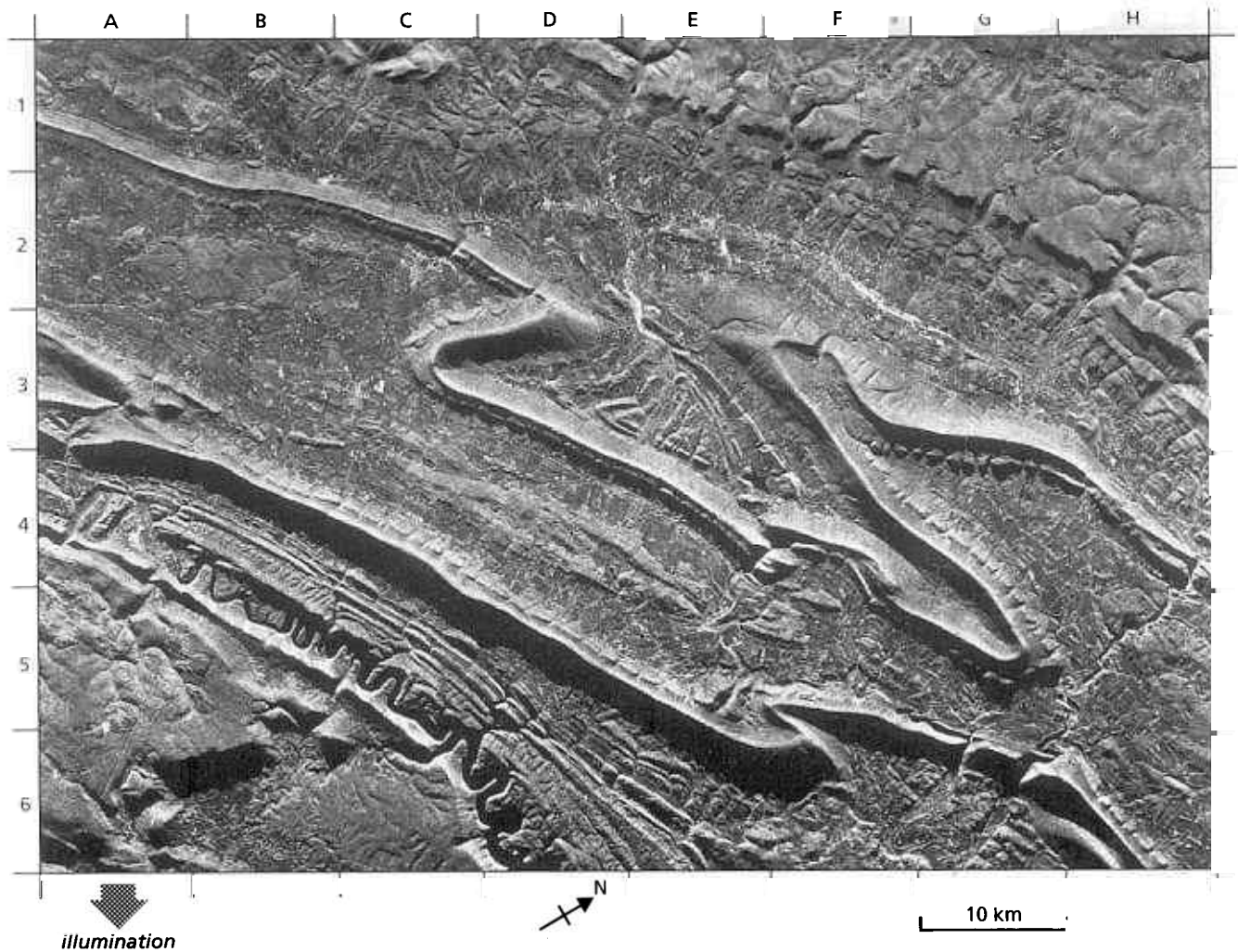


Fig. 7.5 This 1 : 35 000 image of an area near Pittsburgh, Pennsylvania, USA, is from an airborne X-band SAR system looking from west to east. It is 50 km wide, has a resolution cell of 6×12 m (azimuth \times range) and is a digitally processed product of the real-time STAR system. Its high quality enables not only mapping of the gross folding patterns in this part of the Appalachians, but shows up many fine details. At D2 and F2 the two folds are clearly anticlines plunging towards each other. In the synclinal zone at E3 the doubly plunging structure has created a triangular interference pattern. The bright ridge at D5 vs to the south-east, suggesting that the band of ridges from A3/4 to E6 are strata dipping steeply to the south-east. The eastern limb of the syncline at F4 has been displaced by a

fault trending south-west-north-east. Displacement of ridges suggests that it has a downthrow to the south-east. The ridges are densely wooded, hence their relatively bright signatures. A power line cut through the forest between A6 and B5 shows up clearly. Back-scatter is dominated by the vegetation canopy, and does not relate to variations in surface roughness. Nevertheless, the wealth of textural information is useful in identifying different rock types. The ridges are well-drained sandstones, and the valleys are floored mainly by shales and carbonates, which are less resistant to weathering and erosion. The impermeable shale units from A3 to E6 and B1 to H3/4 are characterized by a much finer drainage network than the more permeable strata. Courtesy of Intera Technologies Ltd, Calgary

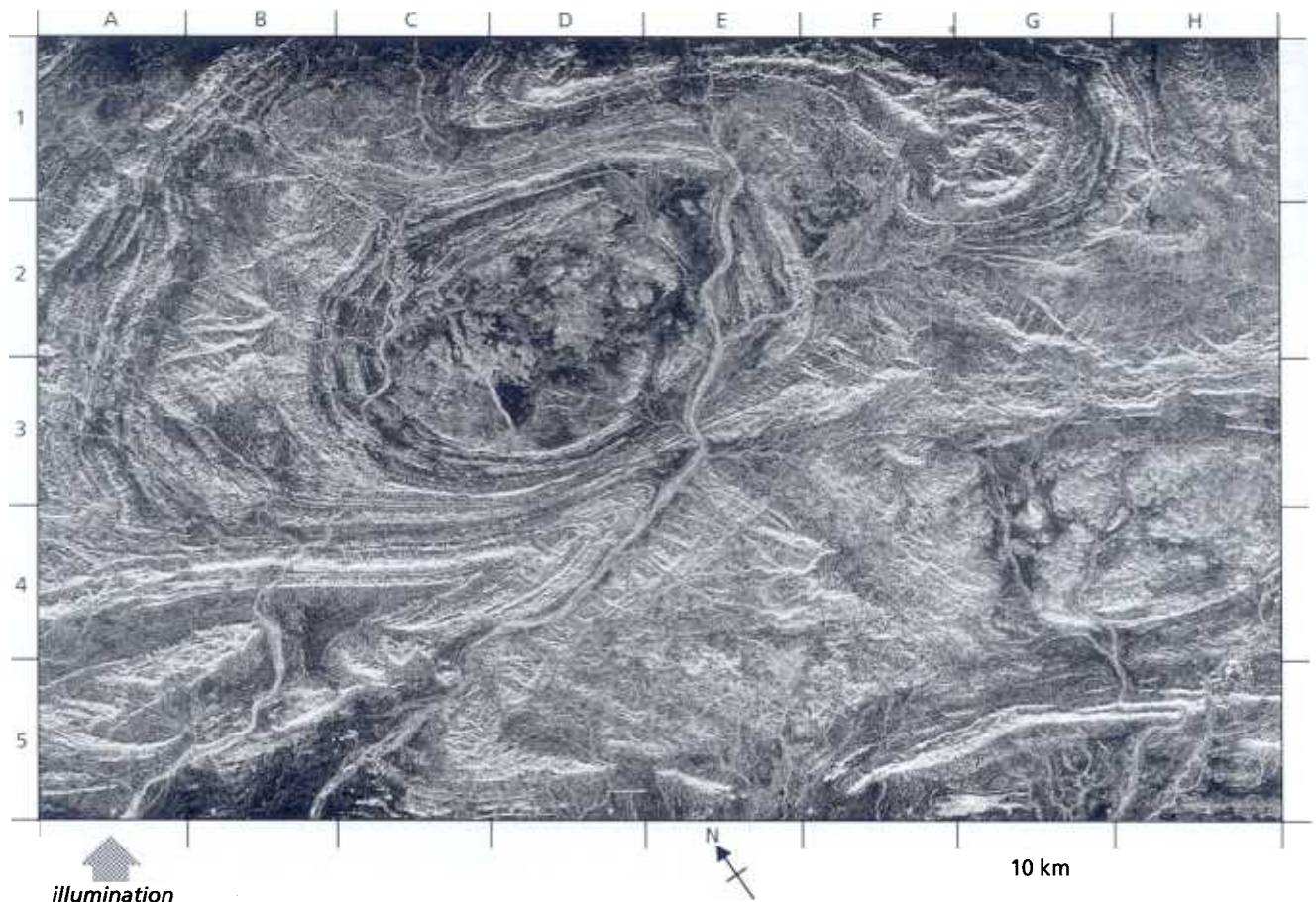


Fig. 7.6 This SIR-A image shows part of the Hammersley Range in Western Australia, at a scale of 1 : 500 000. The area is underlain by Lower Proterozoic volcanics, which are interbedded with some of the world's largest iron ore deposits, in the form of banded ironstone formations. The largest structure is the Rocklea Dome (C2 to E2 and C3 to E3) with a core of Archaean granite. A similar basement-cored dome occurs at H3/4. The bright linear feature from D3 to C2 is a dyke cutting the granite, but which is unconformably overlain by the Proterozoic volcanics. Other fine lines in the granite are also dykes. The dark unit with bright ridges surrounding the dome is an easily weathered sandstone. Succeeding it to the west and between D2, E5 and F4 are pillowed metabasalts, the rough surface of which results in a bright signature. The synform with an axial trace from D4 to

A5 contains a layered sequence of ironstones, shales and carbonates. The radar-smooth unit in the synform consists mainly of shales veneered by alluvium. From D1 to G1/2 is a rugged area underlain by a massive ironstone, weathering of which has produced the huge iron ore deposits of Mount Turner (D1) and Mount Tom Price, which is just north of H1. As well as defining the major fold structures and lithological divisions magnificently, SIR-A also shows up many hitherto unmapped geological features. The most obvious form a series of wide bright linear features in the area from C1 to B3. These are thick quartz veins, many of which follow minor faults. The image fails to detect pisolitic Tertiary limonite and hematite deposits that form the main iron ores, except near Mount Turner (D1), where they show as a bright patch.

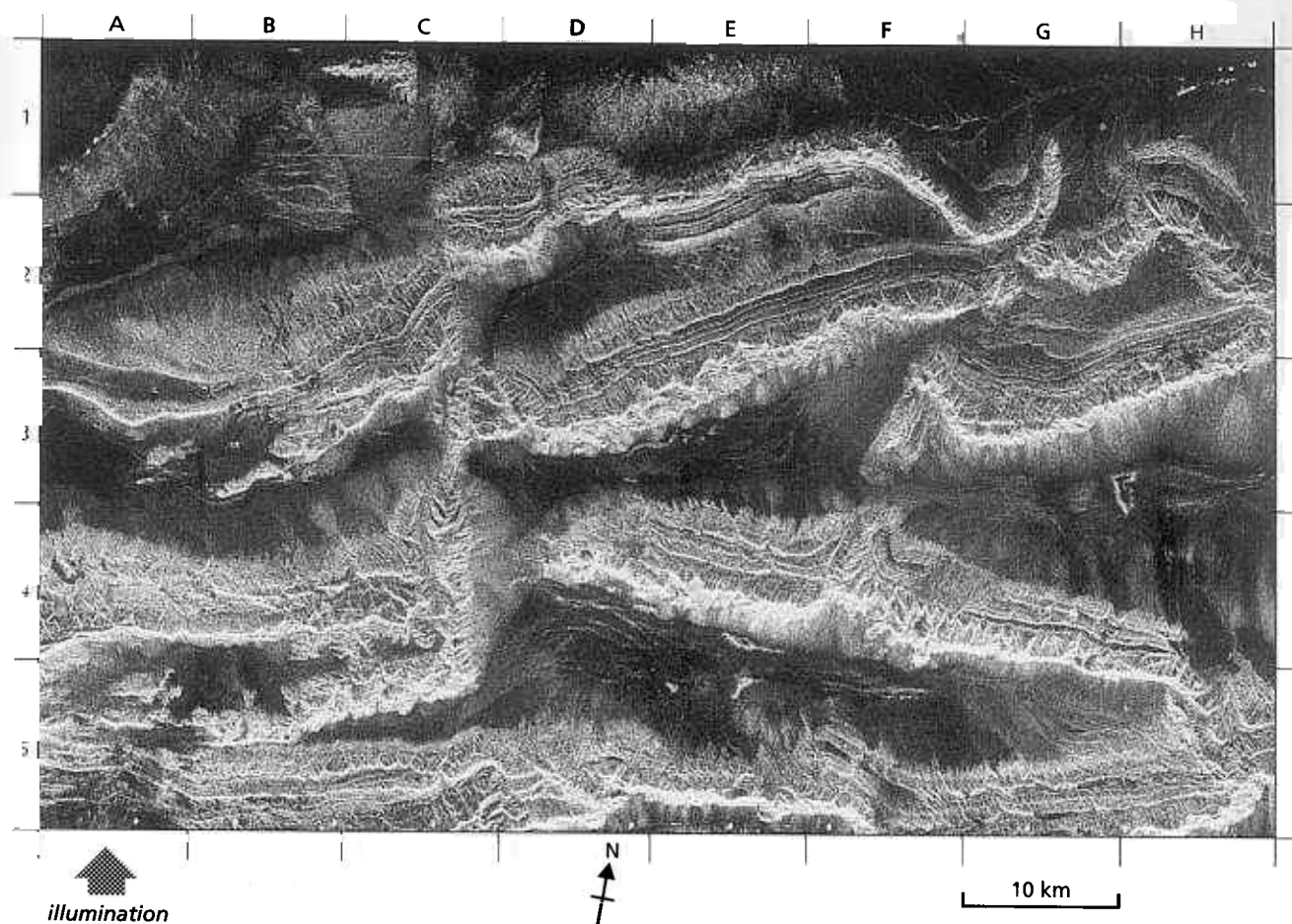


Fig. 7.7 One of the most striking SIR-A images is of the Kalpin Tagh mountains in north-western Xinjiang, China. The area comprises Lower Palaeozoic carbonates and clastic sediments accreted to Asia by plate tectonic events in the Triassic Period. Their structures were reactivated and disrupted during the Cretaceous to Eocene collision of India with Asia, which formed the Himalayas far to the south. The rugged terrain has resulted in some layover. The best displayed of the older structures is a thrust from C2 to G2, which is expressed as a bright line representing a slope angled steeply to the radar beam. It transgresses the layering to the south, especially near the southward lobe at G2, and dips to the north. Determining the dip in the layered strata is difficult because of the prominence of layover flatirons (B/C5, F/G4). Clear Vs in escarpments at E4 and E/F2 show that the two blocks dip north and may be separated by another east-west thrust. Dips

in A/B/C4 are northward and are to the south in A/B/C5, indicating an east-west striking antiform. The east-west structure from G2 to H2 is a synform. The difficulty in assessing the older structures is magnified by the effects of later faulting. The linear feature from C1 to C5 clearly displaces structures on either side in a left-lateral sense and is a major strike-slip fault. Another important fault trends from F4 to G2. In both cases the faulting also affects drainage and alluvium, and obviously has been active recently. An earthquake of magnitude 6.8 on the Richter scale occurred in the area in 1961. In some places (F/G4) huge rock slides, possibly induced by seismicity, have obliterated the layering in the mountains. In the alluvial deposits tone decreases towards the valley axes as granularity decreases, and the distributary gullies are well defined.

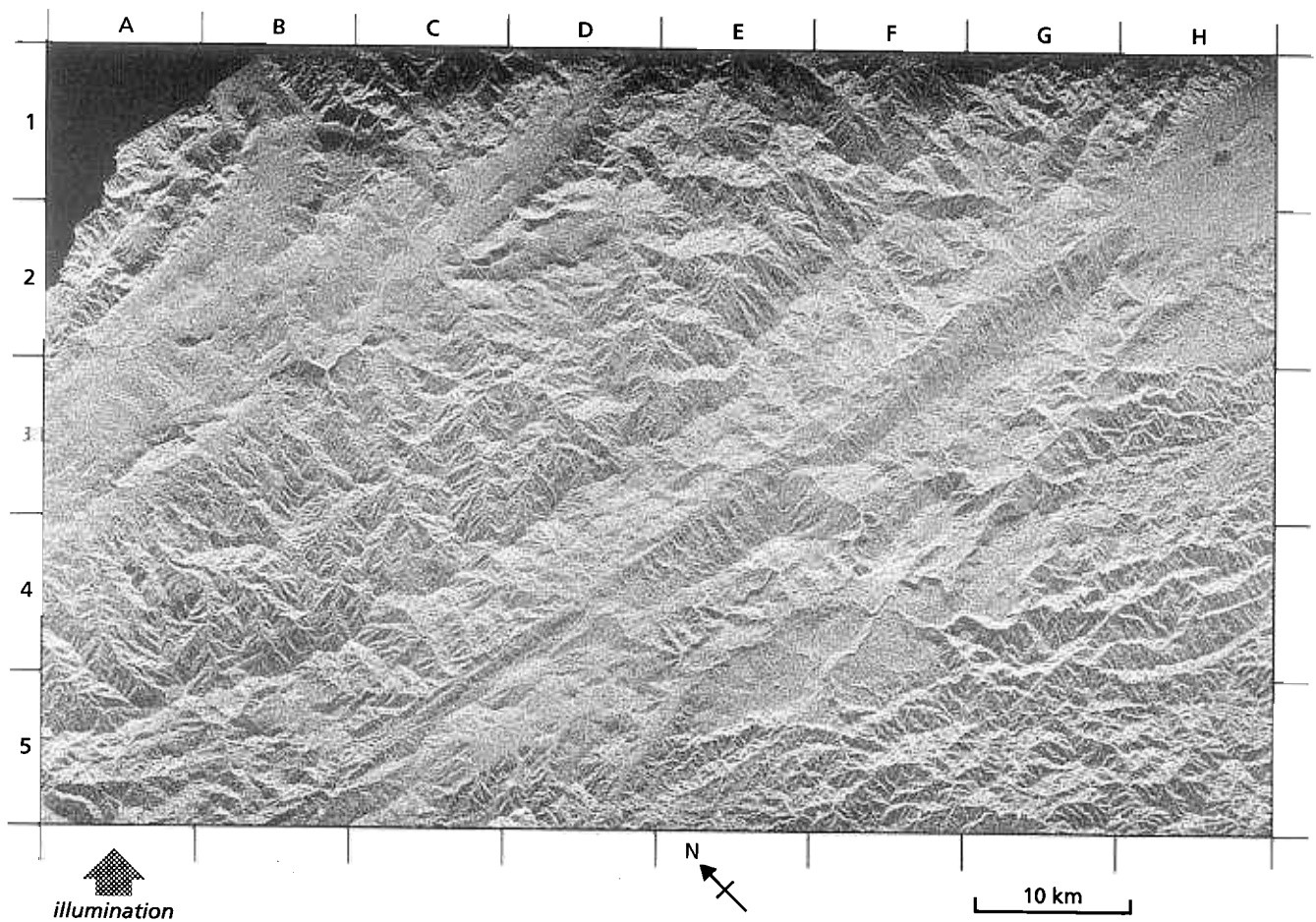


Fig. 7.8 In densely forested tropical terrains, such as this part of the Vogelkop Region in western Irian Jaya, field mapping is hindered by the terrain. Aerial photography and other means of gathering data in the visible and infrared spectra are thwarted by near-continuous cloud cover. Although radar provides useful images, like this one from SIR-A, they are not nearly so sharp as those over arid terrains, because the vegetation canopy intercepts a high proportion of the energy and only a muted representation of the ground surface emerges. However, given experience it is possible to recognize many geological features. The most prominent here is a linear feature trending roughly east-west from H1 to B5. Careful examination of the feature around D/E4 reveals truncated

valleys, deflected drainage and shutter ridges, all reminiscent of an active fault. Its sense of movement is left-lateral strike-slip. Another prominent, possibly active fault trends from H2 to E/F2, and many others can be picked out, especially if the image is viewed obliquely from several directions. Oblique viewing also helps to distinguish different surface textures related to the bedrock. Three units can be distinguished easily. The dissected terrain (H5) is underlain by a Lower Palaeozoic metamorphic complex. The flat surface at E/F4/5 is almost devoid of drainage and probably is underlain by limestone. Along the major fault is a rugged bright area of thrust blocks, but to the north and south are dark, smoother terrains that have developed on sandstones.

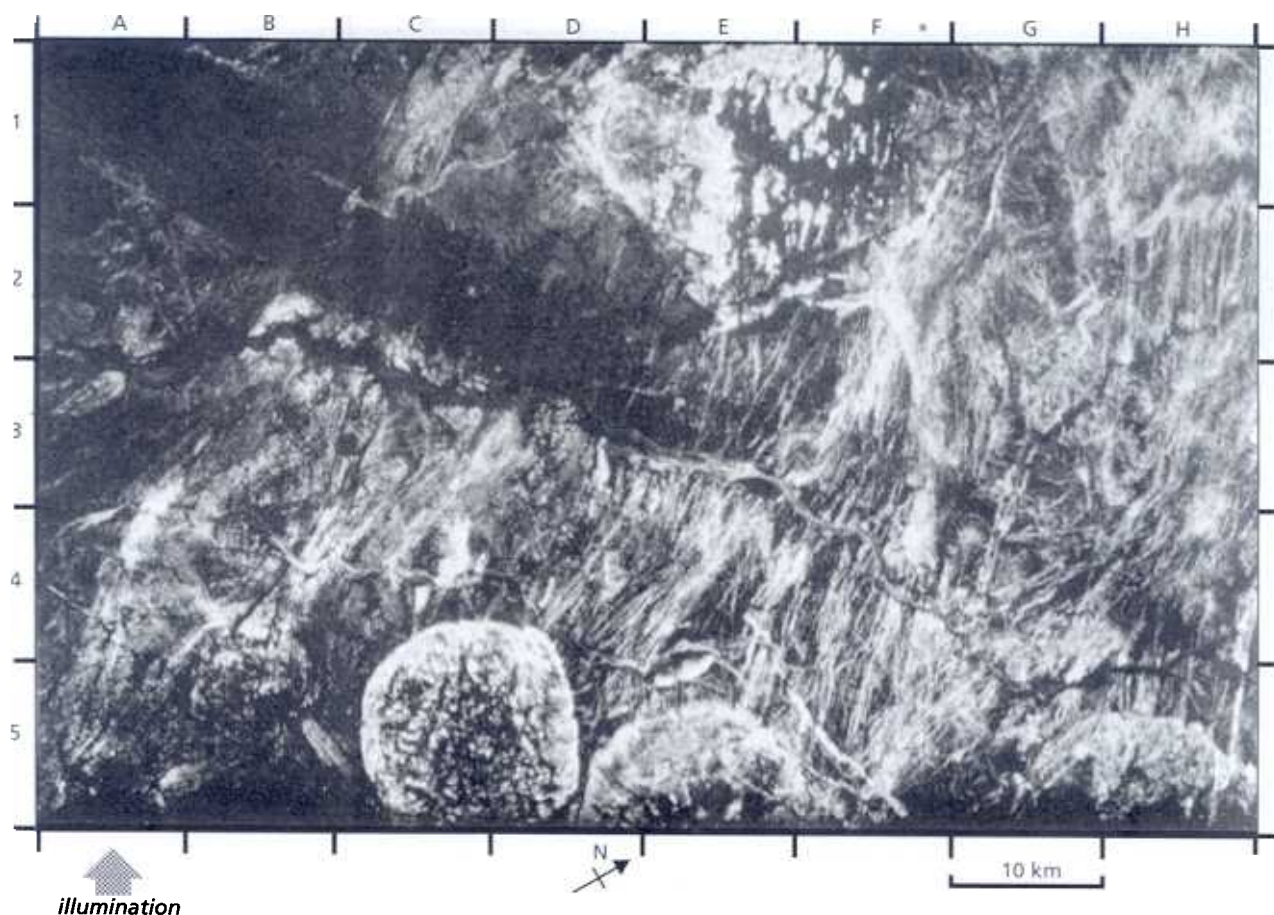


Fig. 7.9 Like Fig. 7.4, this SIR-A image of part of eastern Mali may have penetrated a thin veneer of sand, the high reflectivity of which swamps Landsat MSS images with large areas of bright tones. The dark area in the top left corner is a sand-filled, possibly fault-controlled valley, which is so smooth that all energy is reflected away. The large circular feature at C5 is a diapiric granite penetrating late Proterozoic metamorphic rocks. Its bright rim and dark interior suggest that its outer zones are more resistant than the now sand-covered core. Other granites occur at D/E5, G/H5 and E1. The last one consists mainly of inselbergs set in a sand and gravel pediment, but the roughly circular shape can be picked out. At its south-western boundary it is truncated by a straight linear feature—probably a fault. The scene is dominated by hundreds of dykes, which show as bright lines. Most are orientated

nearly parallel to the look direction, and are at least partly buried by sand. The fact that they show so clearly suggests that they consist of aligned, joint-bounded blocks buried by sand, each of which acts as a corner reflector. Careful mapping shows a complex pattern with a variety of orientations, including an arcuate set at F2. The granites do not seem to have been cut by the dykes and may be later intrusions. The dykes are useful in identifying less obvious granite bodies. At A/B4/5 a swarm of about 10 east-west dykes are not seen in an area bounded by a vaguely circular set of features. This could be another granite. The dominance of dykes in the image obscures most details of the older metamorphic complex that they cut, probably because the older rocks have radar-smooth surfaces. Some curvilinear features around B/C3/4 may define folded banding in the basement.

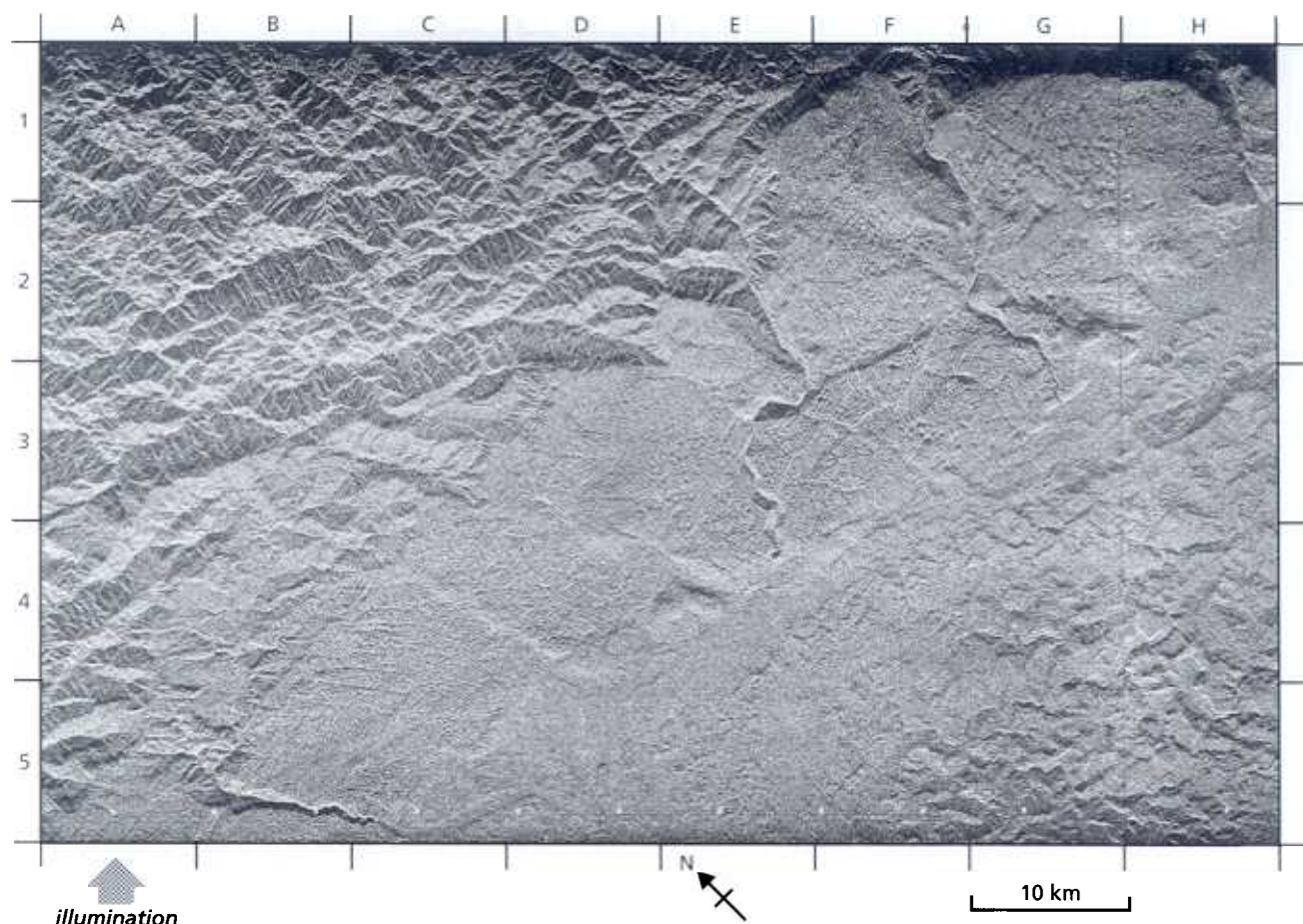


Fig. 7.10 This SIR-A image begins just to the south-east of Fig. 7.8. The top left quadrant is dominated by rugged terrain developed on metamorphic rocks. They are overlain by a small tongue of clastic rocks trending north-west-south-east from A4. The clastic rocks are less rugged, and they have a coarser drainage texture. They are succeeded to the south-east, beyond a line roughly from B/C5 to H1, by rocks on which there is virtually no drainage, and a surface that is pitted with large depressions. They are particularly noticeable in the area G/H1/2. Surfaces like these are typical of tropical karst developed on carbonates. Only three major streams drain the area in deep gorges, and most rainfall, of which Irian Jaya has no shortage, disappears into sink-holes. A low NNW-SSE

scarp between D4 and E4/5 possibly defines a major fault through the carbonates. Indeed, a whole series of muted linear features are visible in this monotonous terrain. A roughly east-west line from D/E5 to H2/3 separates the carbonates from a subtly different kind of landscape to the south. Along the boundary can be seen several small scarps with clear V-shapes (G/H3/4). These are clastic sediments, which dip to the south. The line of Vs is, in fact, a zone of locally steepened dip (a monocline) flanking a major sedimentary basin to the south. Within the basin the drainage becomes clear once more because the rocks that fill it are impermeable.

response from the cell that contains it and thereby show up on an image. Sometimes the back-scatter swamps several cells to produce a star-like bloom. Aligned corners resulting from joint blocks along a small fault or dyke help highlight a linear feature that would otherwise be invisible on an image of reflected short wavelength radiation (Fig. 7.9).

Topographic sharpening also magnifies the contrast associated with smaller scale features, such as drainage patterns and other textural attributes of the surface. They become easier to distinguish and to interpret in lithological terms. Figures 7.8 and 7.10 show examples in

tropical rain forest of various rock-related textures that radar has enhanced.

A further result of the highlighting effect of radar is to make shapes that characterize erosional or constructive processes more distinctive. This is particularly useful in glacial terrains, where strong reflection from snow and ice saturates images of shorter wavelengths. Radar images show variations in surface roughness and therefore ranges of grey over snow and ice. The topography shows quite clearly (Fig. 7.11). Figure 7.12 picks out volcanic cones and the effects of dissection quite well. Much smaller and more intricate drainage channels

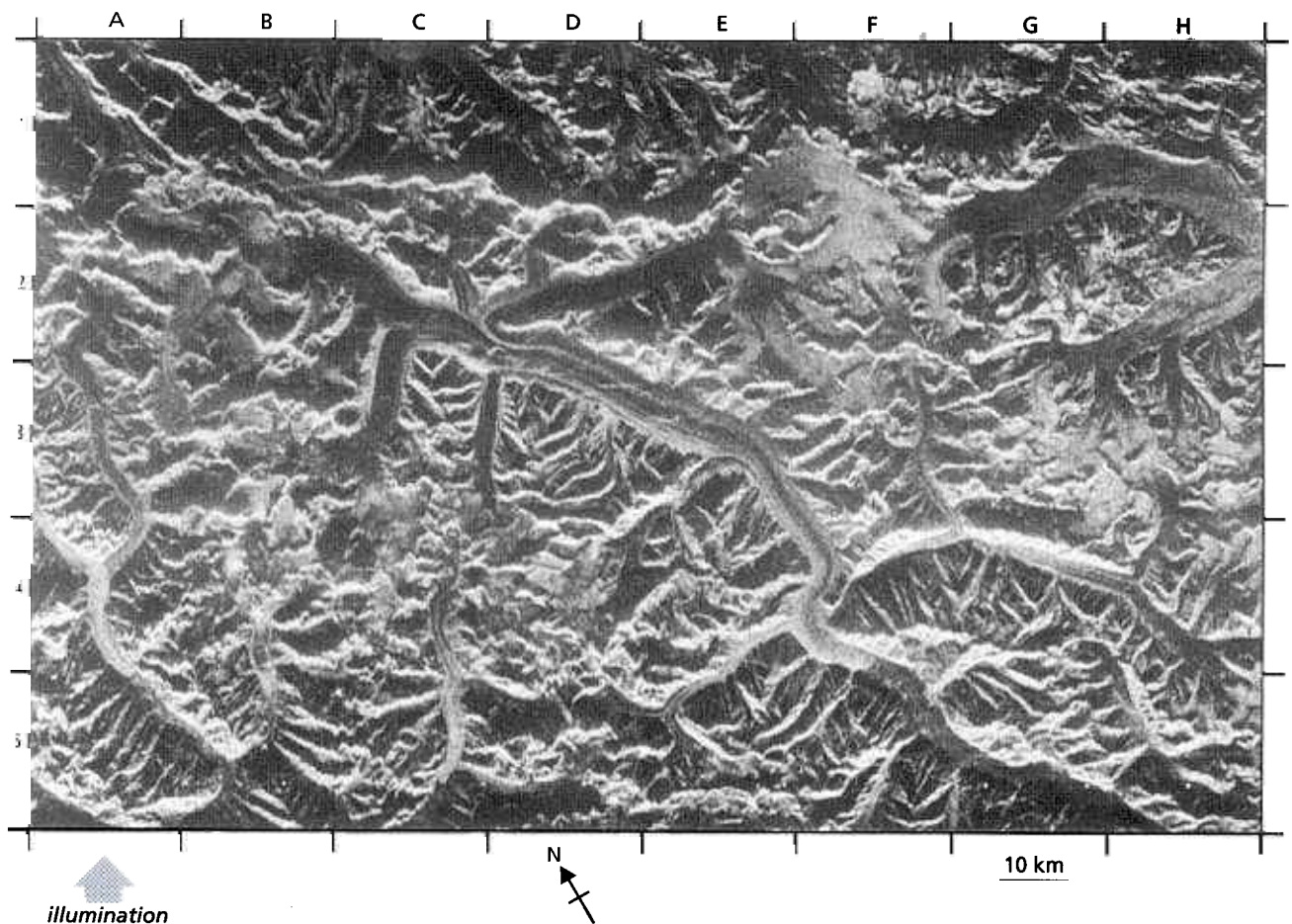


Fig. 7.11 The knot of mountains at the juncture of the Pamir, Salt and Karakoram Ranges formed during the indentation of Tethyan marine sediments, island arcs and microcontinents when India finally collided with Asia some 90–70 Ma. Together they form one of the most rugged terrains on Earth, with the most active erosion known. This SIR-A image of part of the Karakoram on the India–Pakistan border is, as a result, dominated by layover and shadow effects. Very few, if any, details relating to geological structure can be seen because of these distortions. However, it contains probably the most spectacular features of mountain glaciation to be seen anywhere. Snow and ice caps at B2/3 and E/F1/3 show as relatively homogeneous grey areas through which the major peaks show as bright, foreshortened slopes facing the antenna and dark shadows on the opposite side. The relatively bright returns from the ice fields indicate moderately rough surfaces. Each ice field forms the source for major valley glaciers several kilometres across. The glacier in the centre is about 50 km long. Dark and light stripes in it correspond to ice and rougher

medial moraines. Each of the moraines extends to the entry points of tributary glaciers. Bright streaks running across the central glacier are probably ice falls riven with crevasses, those normal to the look direction being preferentially highlighted. The snout of the main glacier, vaguely defined as a down-valley lobe, is at E/F5. Below it the valley has a dark signature, suggesting that it is flat-bottomed. Towards the snout, the glacier becomes brighter and more uniform in tone as melting and ablation reveal more debris and disrupt the pattern of medial moraines. Two tributary glaciers at C3 and D3 appear anomalous, with a uniform black signature. This probably results from their sloping steeply away from the antenna and receiving very little incident radar energy. Although distorted, the radar image shows far more topographic detail than images of other parts of the spectrum do. They would be overwhelmed by the uniform reflectivity and low temperatures of snow and ice, which formed a near total cover in the area at the time of the SIR-A overpass (November 1981).

show up on radar images than on those from other systems (Fig. 7.13). Figure 7.14 shows how radar images of sandy desert landforms are often disappointing because of the general smoothness of sand and its specular reflection of radar energy.

Some terrains present problems for geological interpretation of images that use reflected and thermal

wavelengths. The worst are those with low relief and monotonous reflectivity, and where natural patterns are masked by regular agricultural fields. Both are common in several parts of the world. Included among them are areas with thin highly reflective sand cover, dark desert varnish, burn scars and thin vegetation. The partial dependence of radar back-scatter on surface roughness

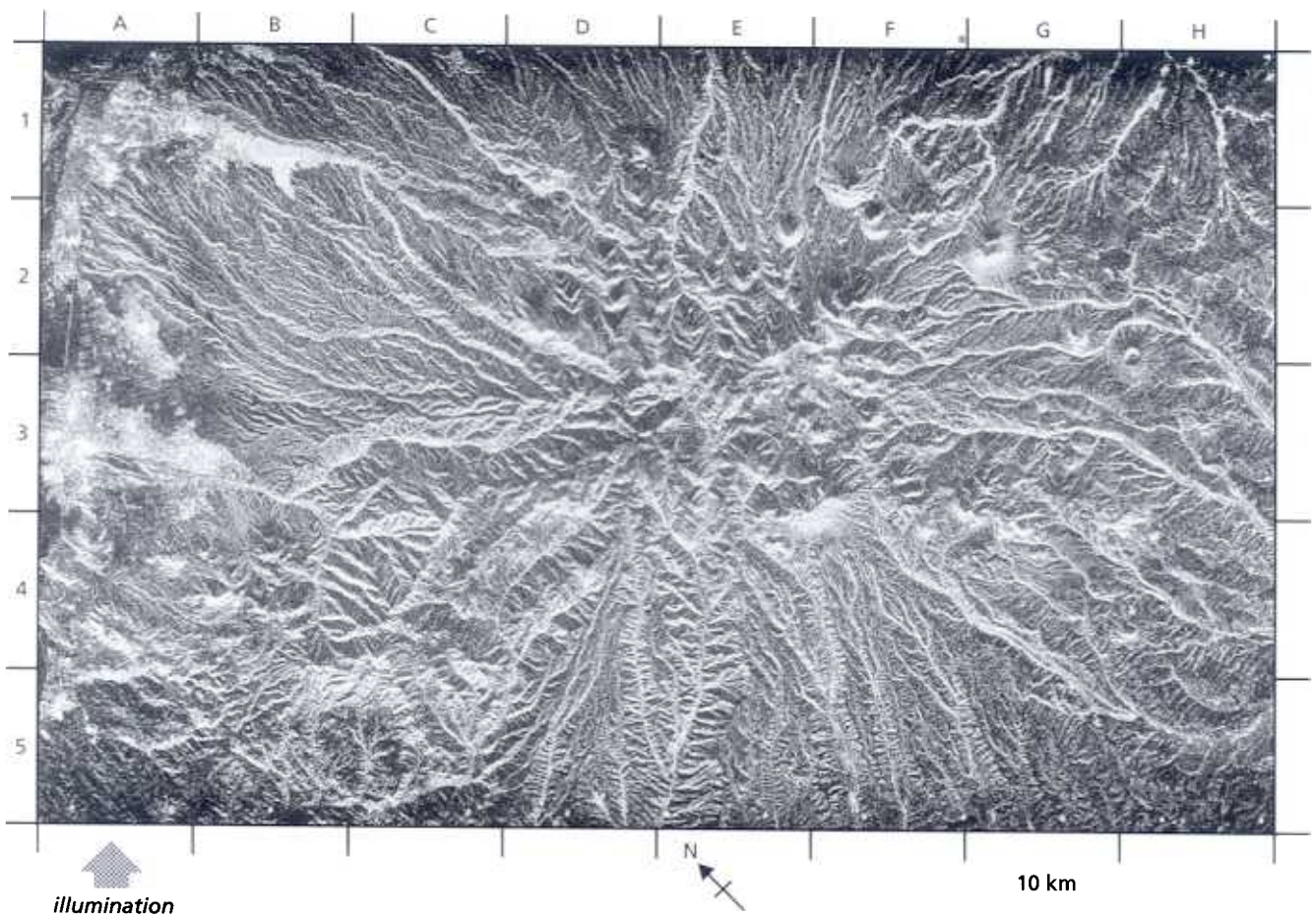


Fig. 7.12 The volcanic edifice of Kuhha-ye Sahand, south of Tabriz in north-western Iran, clearly reveals its structure on this SIR-A image. The marked radial drainage is a common feature of dissected andesitic stratovolcanoes. The valleys radiate outwards from what was a central caldera complex at D/E3, but which is now barely discernible because of deep erosion. Each major valley has an intricate dendritic system of tributaries, in which no annular element can be seen. This suggests a relatively homogeneous structure, perhaps dominated by pyroclastic rocks. Thick lava units separated

by more easily eroded tuffs could tend to impart a concentric element to the drainage. Several younger, parasitic volcanoes can be seen on the flanks of the major edifice from E2 to H3, and around the central zone. Each is clearly a cone, and one has preserved a summit crater (H2/3). On the flanks of these small volcanoes dissection has also begun to impart a radial drainage pattern. In several cases the development of parasitic cones appears to have deflected drainage on the flanks of the main volcano.

allows normally invisible differences in microrelief and vegetation cover to show through such optical masks. In many cases, patterns of great geological significance are revealed as the following examples show.

Basaltic lavas are dark in the visible and near-infrared regions. Radar potentially is capable of discriminating rough aa flows from smooth pahoehoe flows, and between flows of different ages for which roughness is variably modified by vegetation, soil cover and degradation (Fig. 7.15). The response to weathering of the rocks from which it formed often controls soil granularity. Likewise, different grain sizes of alluvium and other superficial deposits relate directly to varying energies of transport and deposition. Chemical sediments, such as salt and gypsum in enclosed evaporating basins can show variations in roughness owing to shrinkage and

localized solution. Such differences show up as tonal patterns on radar images (see Fig. 7.20). They are added to by the changes in dielectric constant induced by different moisture contents. In flat-lying areas, therefore, the tonal variations on radar images can be used directly for geological mapping (Fig. 7.16). However, in more rugged terrains the same effects, though present, are modified by the influence of slope. The radar signature of a uniform surface type changes with the angle and direction of slope and therefore incidence angle (Fig. 7.3). If slopes are chosen carefully, it is possible to see through this and map lithologies with a fair measure of accuracy.

Under conditions of extreme dryness, a proportion of the incoming radar energy penetrates soils. Dryness also permits its return to the surface after subsurface

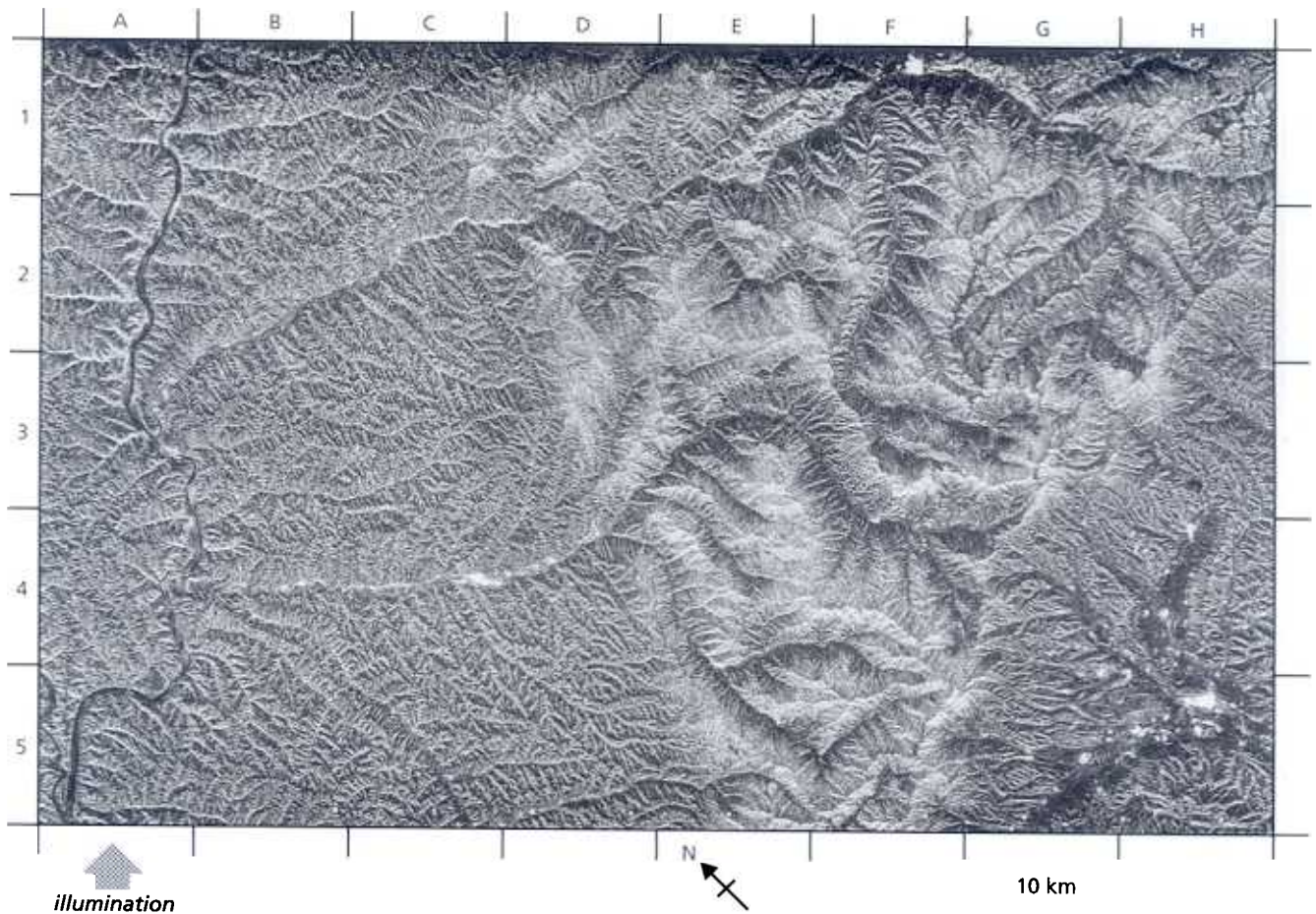


Fig. 7.13 The Huang He (Yellow River) in Shanxi Province, China (A1 to A5) carries a high load of suspended sediment, owing to its catchment comprising large areas mantled by wind blown dust, or loess. This SIR-A image illustrates the complex dendritic patterns of the innumerable tributaries of the Huang He flowing westwards from the mountains in the centre of the image. The changes in structure of drainage from the Hei-ch'a Shan and Luliang Shan mountains to the lower terrains on either side mark major lithological boundaries. The mountains are complex Palaeozoic basement unconformably overlain by Triassic sediments in the lowlands to the east and faulted against the Triassic rocks to the west. The brighter returns from the mountains are to the result of volume scattering from the dense forest canopy above 1500 m. The

lowlands have sparse vegetation except in the valleys, and they show typical badland topography. Many of the smaller tributaries to the west of the mountains display a pinnate drainage pattern. This is typical of areas mantled by thick, homogeneous loess deposits. In some areas this mantle remains, but much of the drainage is probably inherited from a now-eroded loess cover. Although little geological structure is immediately apparent in the lowlands, there is a hint of a roughly north-south grain immediately west of the mountains. Several major valleys have nearly straight courses, which may reflect faults cutting the Triassic sediments (A2, A4 to D4). The bright areas from A4 to D4 and round H4 are major towns in this otherwise sparsely populated region.

scattering. In hyperarid deserts this means that radar images sometimes give direct information about buried features. Radar is the only remote-sensing method where this is possible, but only field studies can confirm that penetration has indeed taken place. A possible example was shown in Fig. 7.4, but in that case the significance of the patterns related only to buried river courses. They probably stemmed from variations in subsurface grain size of the sand. Of more importance to the geologist is energy scattered back to the surface from the interface between sand and bedrock. This allows the topography of the buried surface to be analysed. Erosional patterns pre-dating the inundation by

sand can reveal the same structural information as bare rock surfaces. The only difference is that energy scattered beneath the surface is attenuated, and the patterns appear relatively dark on radar images (Fig. 7.17). As radar refraction at the surface reduces the incidence angle and wavelength, however, the back-scatter sometimes can be increased as more of the subsurface satisfies the roughness criterion (Equation 7.2).

7.2.2 Varying look direction

The position of the Sun restricts the illumination directions that other remote-sensing systems can exploit.

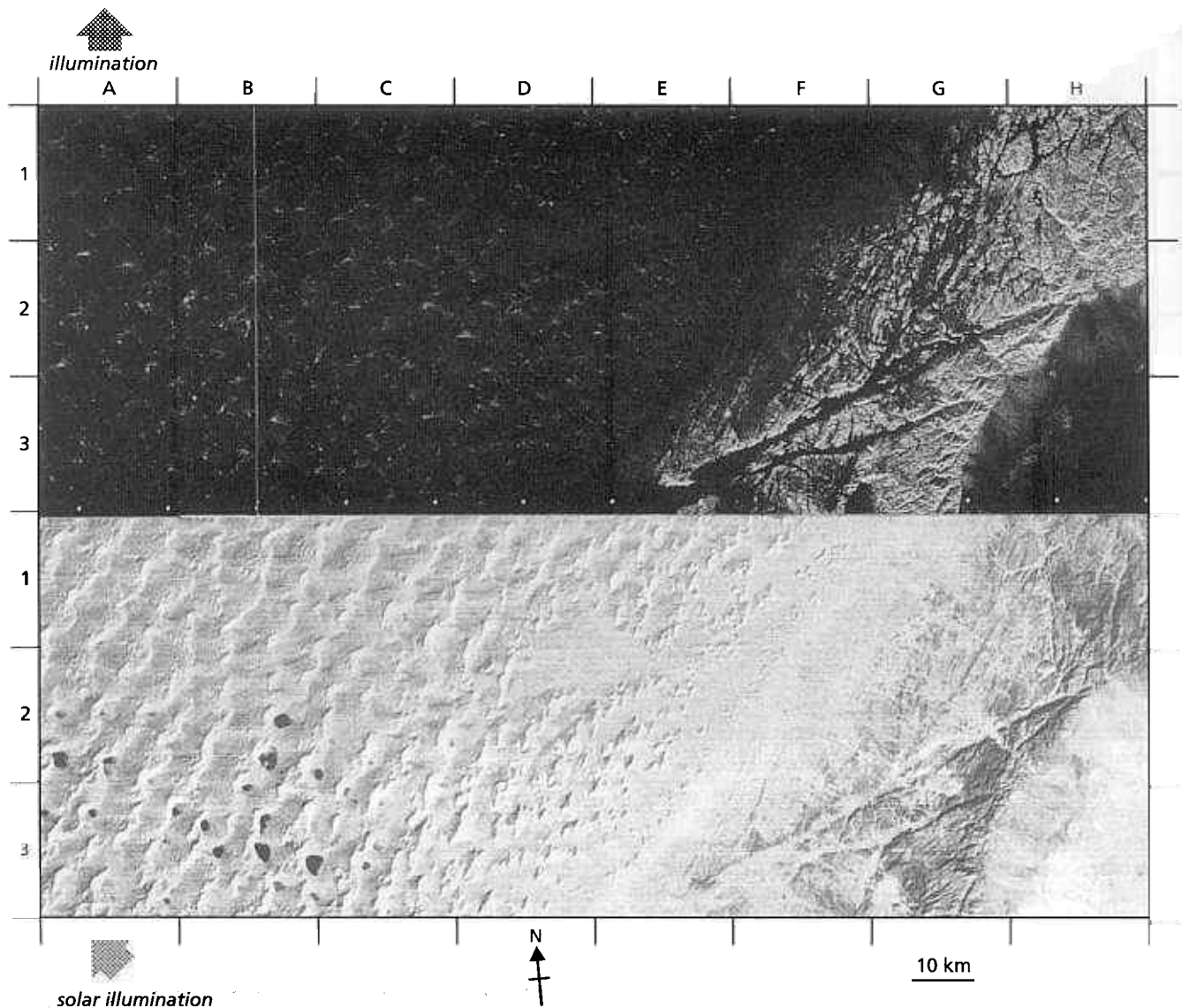


Fig. 7.14 The top of this figure is part of a SIR-A image of the Badain Jaran desert in Inner Mongolia, and the bottom shows a Landsat MSS band 5 image of the same area. The comparison is quite instructive. The Badain Jaran contains the world's largest sand dunes, up to 300 m high. On the Landsat image they show clearly, together with smaller dunes superimposed on them. In the west they are huge composite dunes with a crescentic or barchan-like component. Towards the mountains they grade into sharp-crested ridges with a north-westerly trend. Some of the interdune depressions contain spring-fed lakes. On the SIR-A image very little information on dune structure is visible. This is because sand is smooth to L-band radar, and only dune slopes facing south return energy to the antenna. The main trend of dunes is parallel to the look direction, so the whole area appears mainly dark. The bright

patches are favourably orientated dune faces and thickets of reeds surrounding the interdune lakes. The low information content of radar images of sand-covered areas is in sharp contrast with the response from areas of outcrop. There, the Landsat image is 'washed out' by highly reflective thin sand cover, whereas the radar image is sharp and informative. This is a result of low returns from deep sand-filling depressions, and high returns from both bare rocks and surfaces with only thin, dry sand cover, which the radar has penetrated. Much more structural information is present in the radar image. Rock surfaces in deserts are generally rough because of wind erosion. At E/F3 on the radar image can be seen evidence for this in the form of narrow ribs of rock trending north-east for about 10 km. These are *yardangs*, or aerodynamic ridges sculpted by sand-laden prevailing winds.

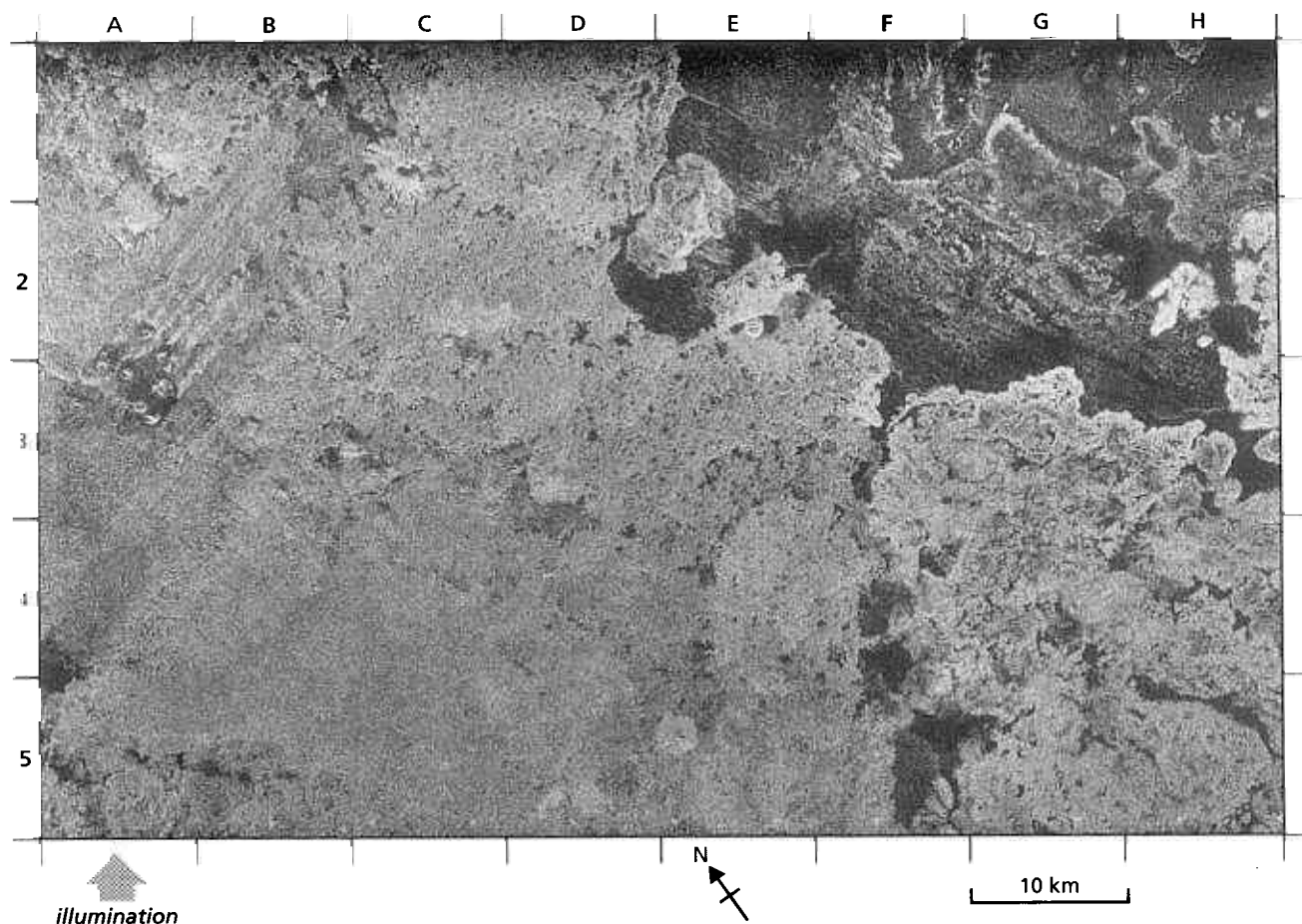


Fig. 7.15 The Jebel al Druz in western Syria is a large Quaternary volcanic field dominated by basaltic lavas, the north-western part of which is shown on this SIR-A image. Much of the surface is extremely rough, so the lava field shows up as medium to light grey-tones. On Landsat MSS images it is very dark. Here the boundary between lavas and the smooth, dark, older surface on to which they flowed is sharply defined. Its lobate form from E1 to H3 is typical of the flow fronts of fluid lavas. Much of the lava surface is speckled with black patches. These are pits and depressions in the flows, now partly filled with wind-blown sand. As a general rule, lava surfaces become smoother with age as surface processes degrade their original morphology, so the youngest lavas should be the brightest on radar images. At D/E5, E2 and C1 are particularly bright surfaces. In each case a small cone is associated with these rough surfaces, and at C1 the area is made up of many tongue-shaped lobes, confirming the

presence of young flows. In the area from A2/3 to C2/3 are a number of small, young cones with vents. Those at A2/3 are surrounded by a dark patch, suggesting that they produced pyroclastic deposits which form a locally smooth surface. Streaks running to the north-east from this area show how the finer ashes have been transported by the prevailing south-westerly wind. To the south-east of this cluster of cones is a north-west-south-east line, which proves to be an array of small cones on closer inspection. These are probably controlled by a fault in the basement. A similar feature extends between A1/2 and C3. These north-west-south-east arrays of minor vents are common throughout the Jebel al Druz, and they are found as far south as southern Jordan. They seem to be connected to faults which splay out from the major left-lateral strike-slip system along the Jordan valley and the Gulf of Aqaba.

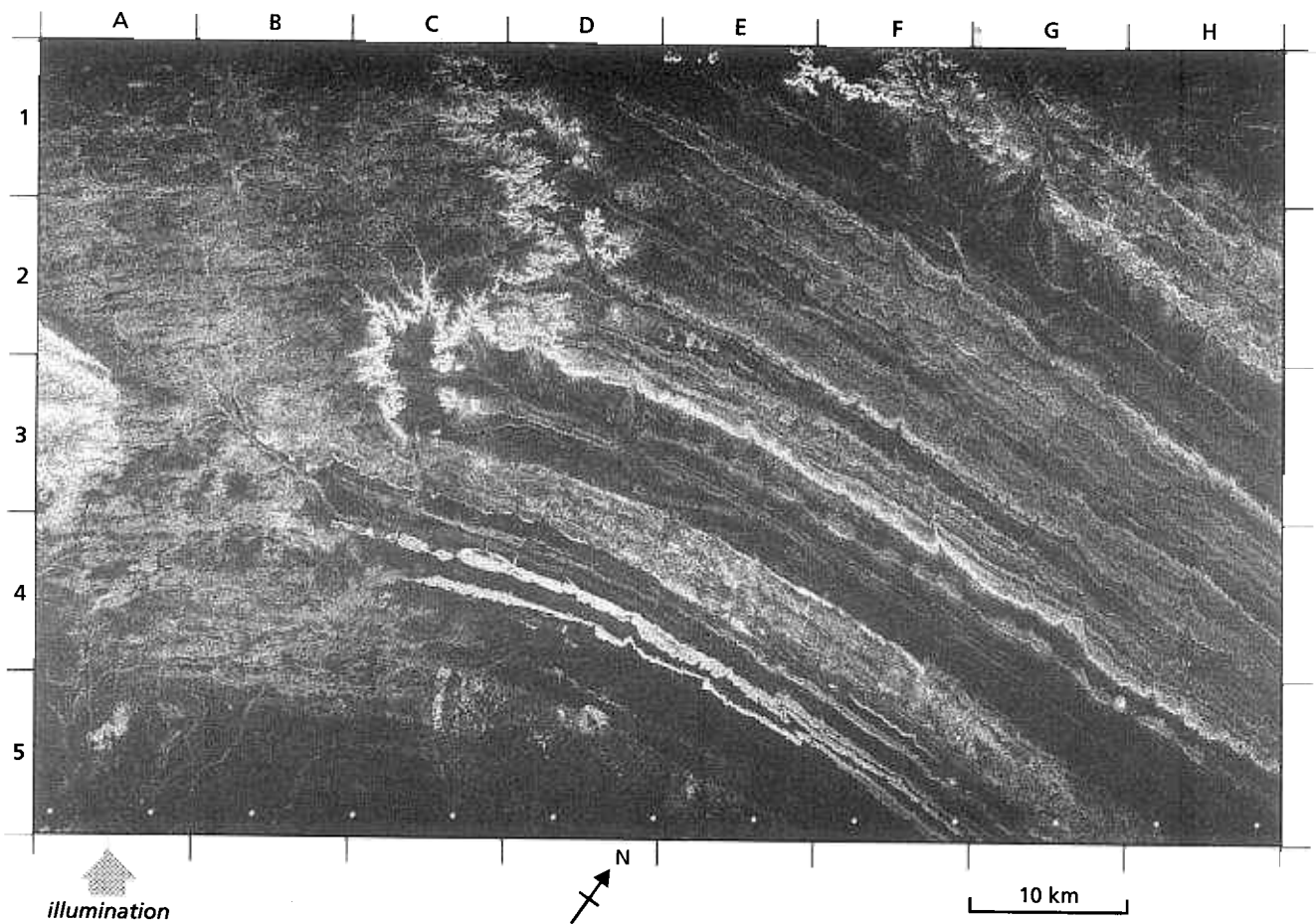
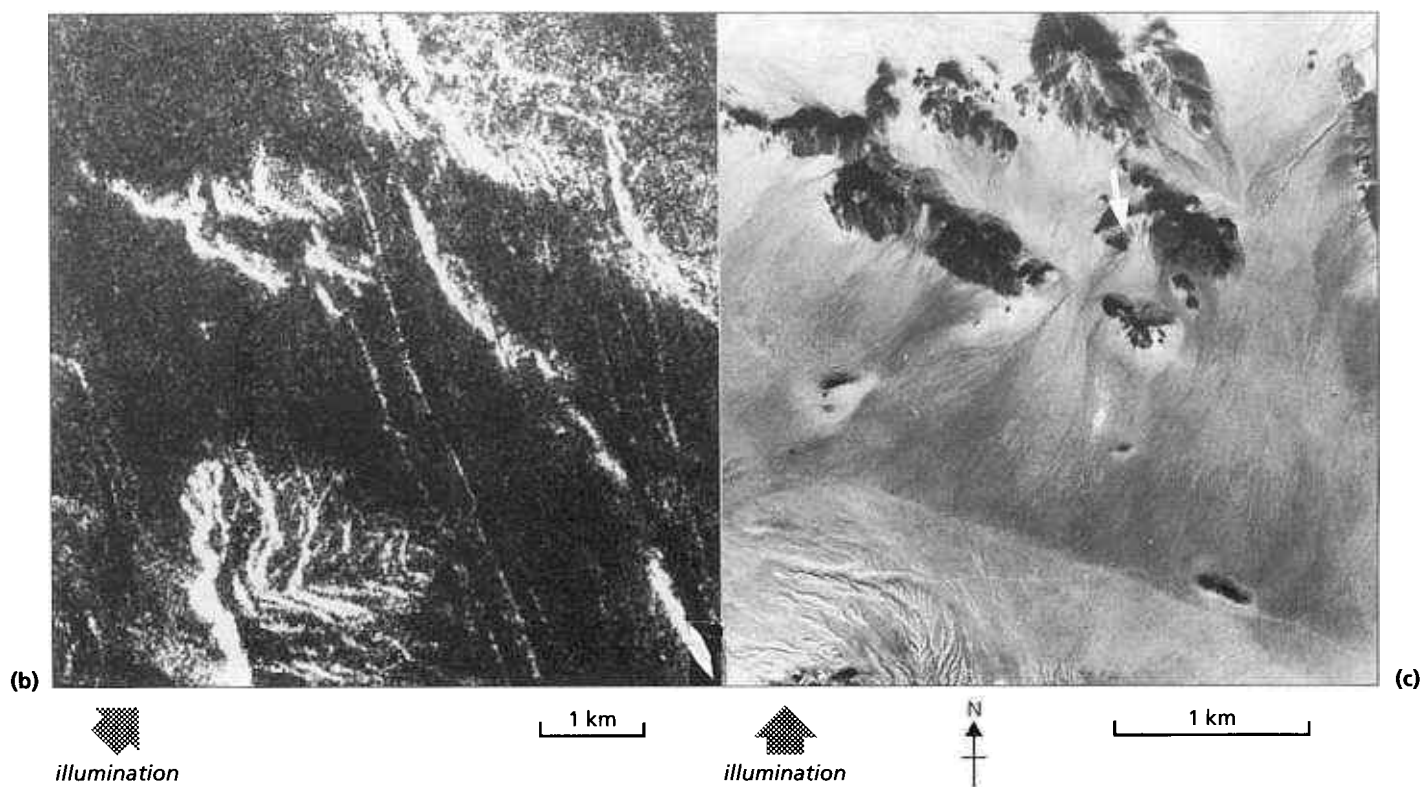
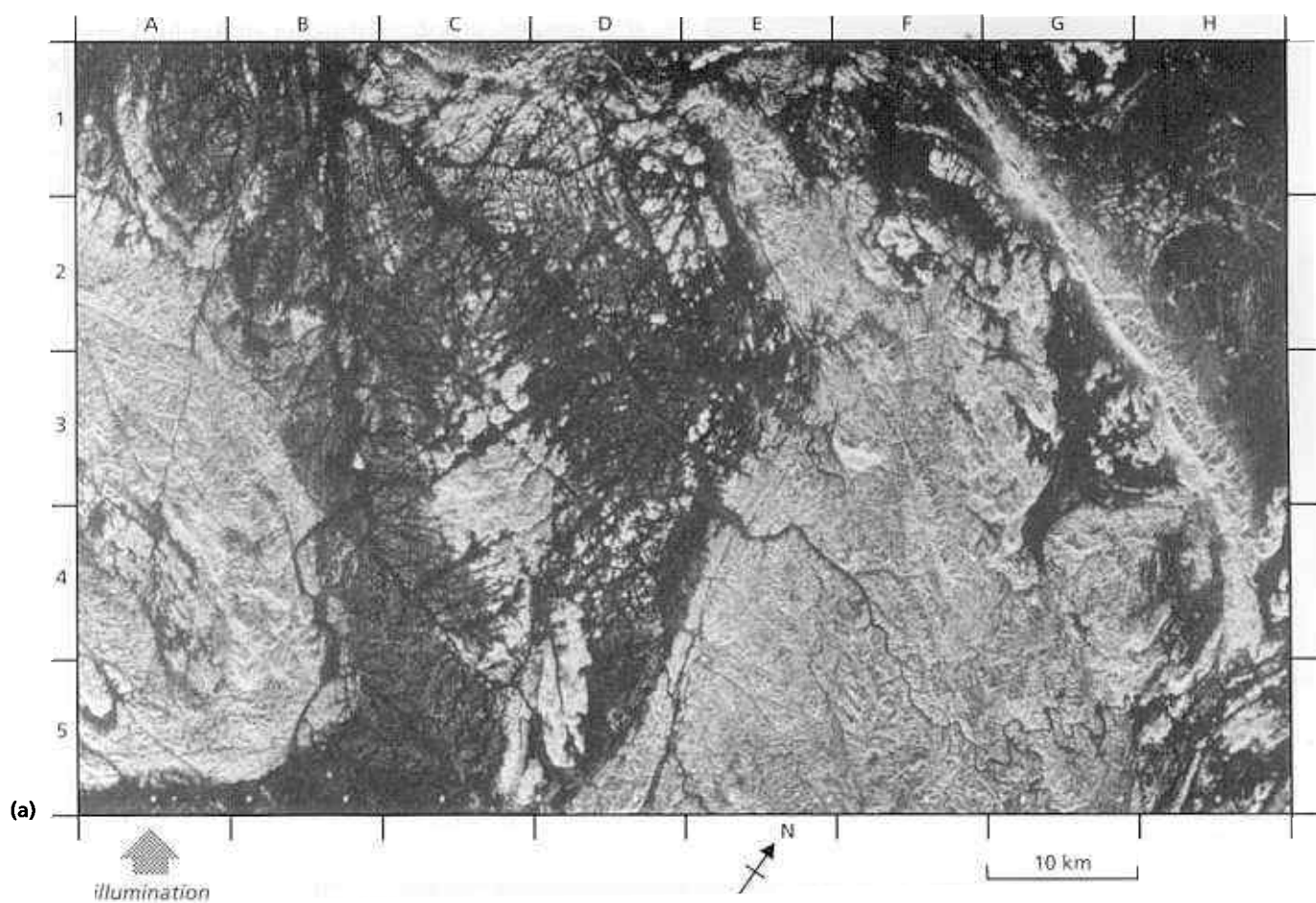


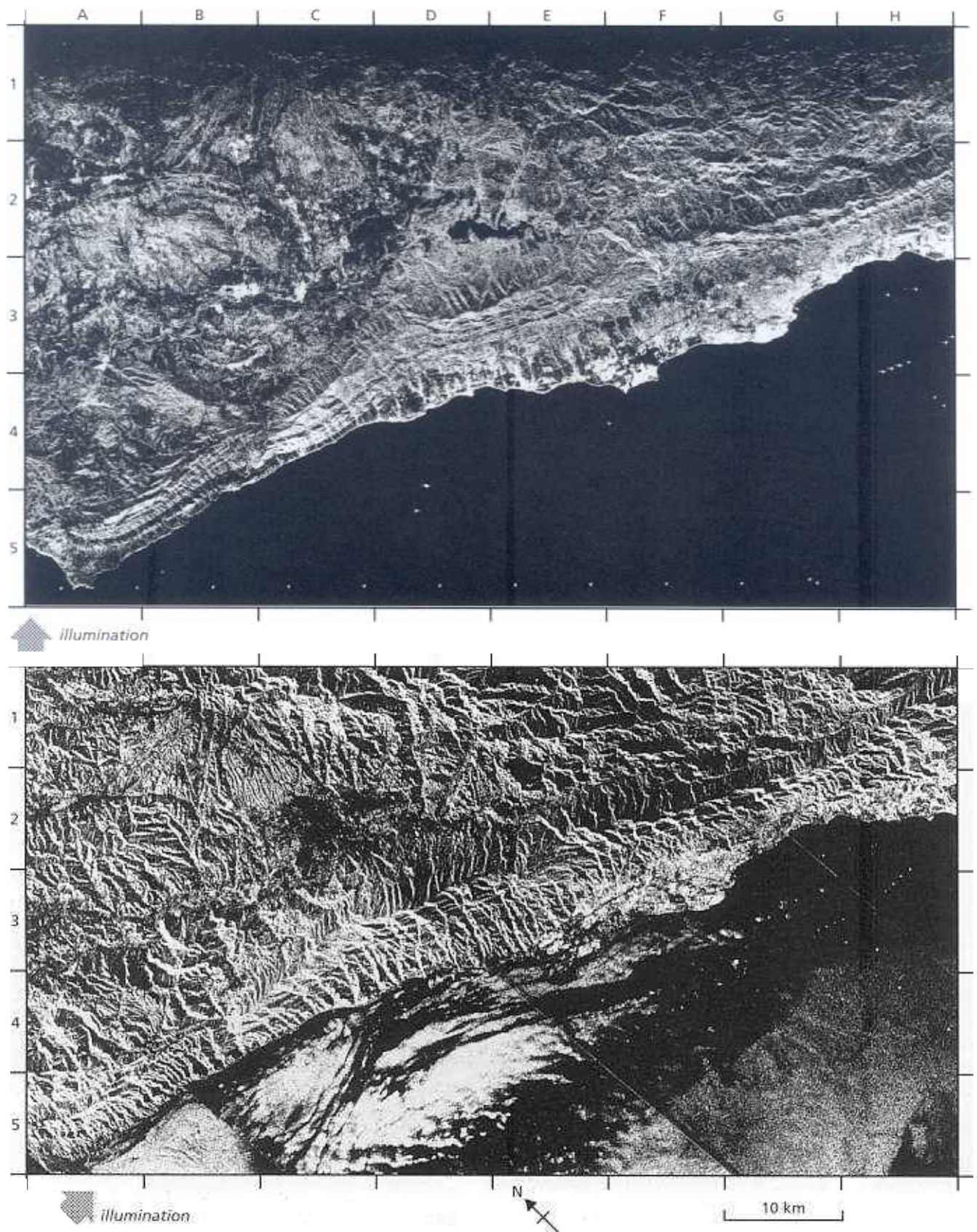
Fig. 7.16 The Devonian sediments of Western Sahara, which are shown on this SIR-A image, are mainly siliciclastic, ranging from conglomerates to siltstones. They are approximately the same in composition, and present similar signatures on images in the visible and near-infrared regions. Here they are magnificently divided into easily mapped units on the basis of their radar back-scattering properties. The desert surface is nearly flat, and the tonal banding represents variations in outcrop roughness. Because individual beds can be mapped, small-scale structures show up well, such as the folds from D4 to E5 and small faults from F2 to F4 and G2 to G5. The presence of V-shaped outcrops controlled by small valleys shows that dip is to the north and the oldest rocks are at C/D5. The bright, irregular boundary from C1 to C3 is an escarpment, facets of which act as corner reflectors to give specular returns. The

escarpment truncates the Devonian bedding and represents a nearly horizontal unconformity, above which are much younger sediments. Their radar signature is monotonous because they are undissected. Lines running south-west-north-east on the surface may be sand streaks parallel to the prevailing wind. Another possibility is that they are related to joints. This would explain the uniformly bright signature of the escarpment, irrespective of its orientation. Joints with this direction would produce many facets on the escarpment that are perpendicular to the radar look direction, and would ensure specular returns. Another interesting feature is that the very few drainage channels in the area have bright signatures. This suggests that they are filled with pebbles and bouldery material. The supply of fine-grained material in such a rocky desert is probably low.

Fig. 7.17 (opposite) (a) The Red Sea Hills, a narrow swath of which is shown by this SIR-A image running south-west-north-east near the Sudan-Egypt border, are composed of a complex assemblage of Precambrian and Lower Palaeozoic metamorphic and igneous rocks. The bright areas are free of sand cover, and they are very rough. They are almost black in the field, owing to desert varnish. The darker areas are mantled by high-albedo sand. As the sand is very dry, radar has penetrated it to show features at the sand-rock interface, but has been attenuated. Large numbers of linear features show up, some of which control subsurface extensions of the wadis that cut the exposed rock surface. As well as aiding geological mapping of the areas, the penetrating capacity of radar provides a unique means of predicting where groundwater may have accumulated in the subsurface channels. Image (b) was produced after digital processing and

enlargement of Seasat radar data for part of the semiarid Mojave Desert in California, USA. Image (c) is an aerial photograph of the left-central part of the area covered by the Seasat image. The radar image shows several NNW-SSE-trending bright lines, whereas the photograph shows no sign of these features, which are in a monotonous area of alluvium. Detailed field mapping, and geophysical surveys across the radar features showed them to be buried quartz-monzonite dykes. That they show up on the Seasat image is the result of four factors: the dryness of the sand cover at the time the image was acquired, the fortuitous orientation of the dykes across the radar look direction, marked subsurface relief on the dykes which caused them to act as elongated corner reflectors and the rubbly nature of the products of dyke weathering compared with those derived from country rock. Courtesy of Ron Blom, Jet Propulsion Laboratory, Pasadena.





Radar is subject to no such hindrance because it is pointable, although orbital parameters restrict this from satellites. Topographic enhancement by low Sun angle is limited to early morning and late afternoon. Satellite systems, such as Landsat, gather data at fixed times of day. This suppresses topographic features roughly orientated along the solar illumination direction. Although the degree of suppression of features parallel to illumination is greater in radar images than with other methods, only two or three carefully planned sorties reveal everything that is detectable. Orbital systems permit four look directions—to either side from both ascending and descending orbits. Selective pointing of the antenna by electronic means can increase the range of directions even further.

There are two main advantages of multiple look directions. First is the selective enhancement of differently orientated topographic features. Careful planning allows the analysis of structural features with all trends and geometries. Moreover, features hidden in shadow from one look direction are revealed from another direction. Figure 7.18 shows examples of SIR-A and Seasat imagery of the same area taken from orbits inclined at about 45° to each other.

Radar's other advantage relates to directional features at the surface that are too small to be resolved, such as parallel corrugations in ploughed fields and on weathered outcrops of finely banded or foliated rocks. Radar looking along the corrugations responds as though the surface is smooth—it appears dark. A look direction across them gives the effect of roughness—the surface appears bright—provided the amplitude and wavelength of the corrugations satisfy the criterion in Equation 7.2. For sand ripples these conditions are not met—they are not sharp enough. Although some research has been directed at detecting hitherto undiscovered potato crops, the potential for geological analysis so far remains untapped. Figure 7.19 shows this effect in a rather unusual agricultural context.

There is one major disadvantage of using different look directions. The appearance of rugged topography is very different on radar images looking in two separate directions (Section 3.5; Fig. 7.18) thereby disrupting interpretation.

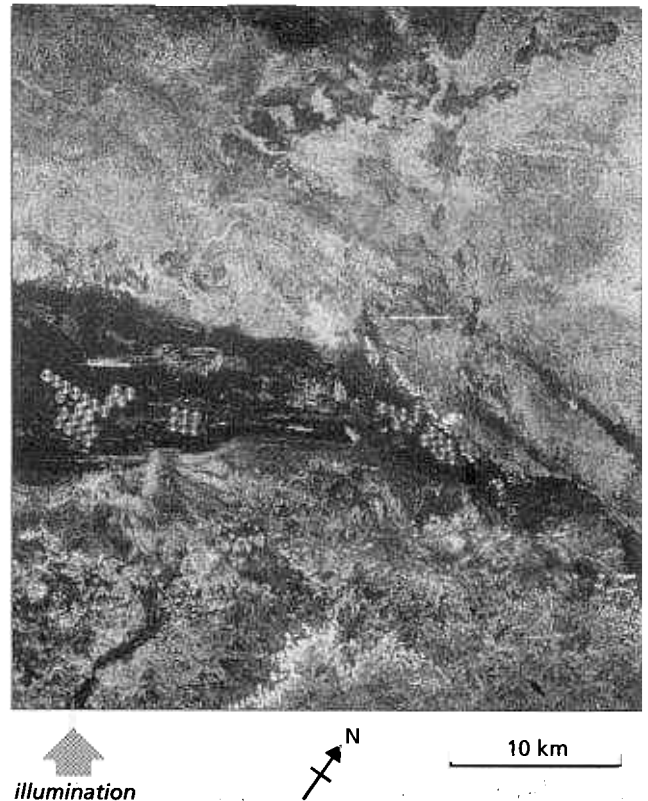


Fig. 7.19 The circular patches on this SIR-A image of northern Libya are fields irrigated by centre-pivot sprinkler systems, supplied by deep boreholes. Only furrow slopes normal to the radar illumination are able to reflect energy back to the sensor. Closer examination shows that the circles have structure shaped like bow ties because the curvature of the furrows becomes less extreme towards the perimeters of the fields. This effect is the result of corrugations far smaller than the image resolution and may also occur on ribbed rock outcrops.

7.2.3 Varying depression angle

Whereas any radar image can give information on surface roughness, radar science presents two possibilities for finer division of this attribute. Section 7.2.5 examines the potential of different radar frequencies. This section concentrates on what may be possible with different radar depression angles. Figure 7.3 implies that steep angles give a maximum response from surfaces that are smooth

Fig. 7.18 (*opposite*) The top image is from SIR-A and the lower from Seasat. The image swaths were gathered from orbits inclined at about 45° to each other. The area is the Santa Ynez Mountains of coastal California, USA, centred on Santa Barbara at E/F4. There are obvious differences between the two images, owing to the different depression angles of the two systems. The Seasat image has stronger layover, and it shows subtle features at the sea surface. However, many of the differences stem from the contrasted look directions. The strike of layered rocks along the coast is shown well by SIR-A, because it is favourably

orientated towards the look direction. This attribute is disrupted on the Seasat image by strong returns from north-south valley sides. Running from D1 to B/C5 is the Santa Ynez Fault, which is clear on the SIR-A image, together with indications of the displacement of bedding. It is nearly undetectable on the lower image. Seasat, however, highlights a major fault from B1 to A/B4 that is not visible on the top image because of its unfavourable orientation. The smaller depression angle of SIR-A results in a much clearer expression of variations in surface roughness. Courtesy of J.P. Ford, Jet Propulsion Laboratory, Pasadena.

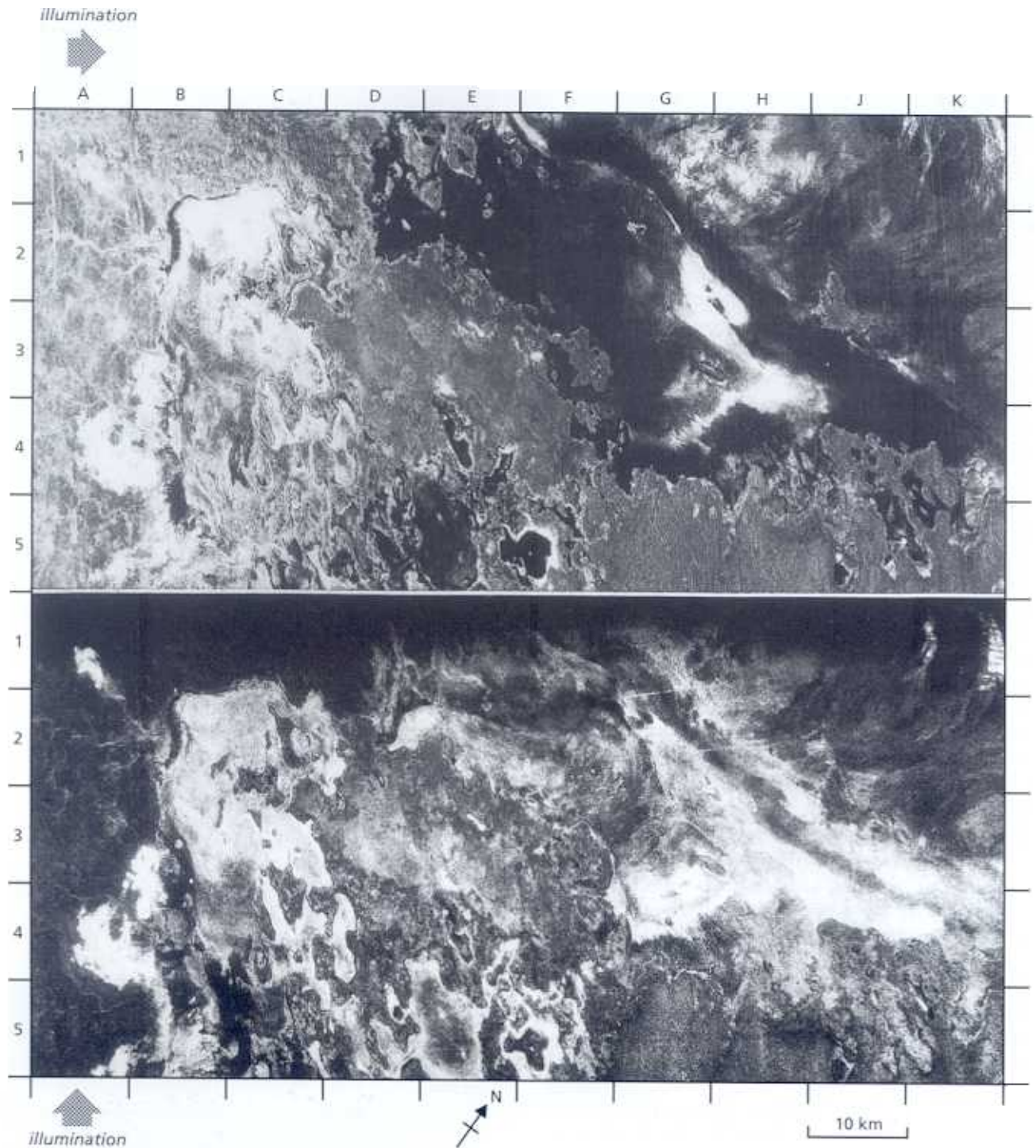


Fig. 7.20 The top image is from Seasat and the lower from SIR-A L-band radar. Although the illumination directions are at right angles, the area has too little relief for this to have any significant effect. The obvious differences between the two images derive mainly from the different depression angles (70° and 43° , respectively). The area is the Sahara Desert of north-eastern Algeria, dominated by evaporite basins, known locally as *chotts*, which are below sea level (B2 to C2 through B5 to C5, and D2 and G2 through F4 to K4). The western chott has roughly the same brightness on both images, suggesting

a moderately rough surface; however, the Seasat image contains more information. In contrast, the eastern chott appears monotonously dark apart from a swirling bright band on the Seasat image, whereas the SIR-A image shows stronger and more variable returns. The differences could result from subtle variations in the roughness of the salt deposits that floor the chotts, and which are variably enhanced by the different depression angles. Courtesy of J. P. Ford, Jet Propulsion Laboratory, Pasadena.

for a particular wavelength, according to the Rayleigh criterion. Shallow angles eliminate back-scatter from all but the roughest surfaces. Angles between 30 and 40° give different responses from all sorts of surface.

Variations in depression angle also accentuate surface topography in different ways because of the different degrees of layover, foreshortening and shadowing (Section 3.5). In particular steep angles maximize the discrimination of subtle changes in areas of low relief. This has to be traded against the distortion of high-relief features, which are better expressed by shallow depression angles. In turn, low-angle illumination may miss subdued features in the landscape.

Using different depression angles is then a means of increasing information potentially related to both lithology and structure. Although most research has been conducted with aircraft data, the most widely available comparisons come from the Seasat and SIR-A orbital missions—both deploying L-band radar with depression angles of 70° and 43° , respectively. Figure 7.20 gives some idea of the potential of systems, such as Radarsat, that permit multiple depression angles.

7.2.4 Stereoscopic radar images

An inherent characteristic of side-looking radar is its inevitable introduction of relief displacement on images (Section 3.5). As it measures the down-range distance of an object in terms of time taken for energy to reach and return from the object, the displacement is towards the nadir of the system. In images produced by optical systems, relief displacement is away from the nadir. Either kind of relief displacement produces parallax between images taken from different positions, and therefore the possibility of stereoptic viewing (Section 2.5, Appendix A). How it is produced makes little difference to the observer.

There are several radar configurations that produce stereoscopic potential (Fig. 7.21). Images with the same depression angle from each side of an area, or those from the same side but with different depression produce stereoscopic convergence.

Depression angle and its geometric effects vary considerably across airborne radar images from near to far range. On a pair of images the conditions for stereoptic fusion and for vertical exaggeration change from place to place. The observer may need to be an ocular athlete. This variation is almost absent from orbital images, so they are potentially much easier to use. The main problems are those of steep slopes and rugged relief, and their influence on layover, foreshortening and shadow effects. Provided such effects all have the same sense, as they do on pairs of images acquired from the same side of an area, information is sensible and can produce photogrammetric quality stereopairs. Images taken with opposed look directions are all but useless in areas of high

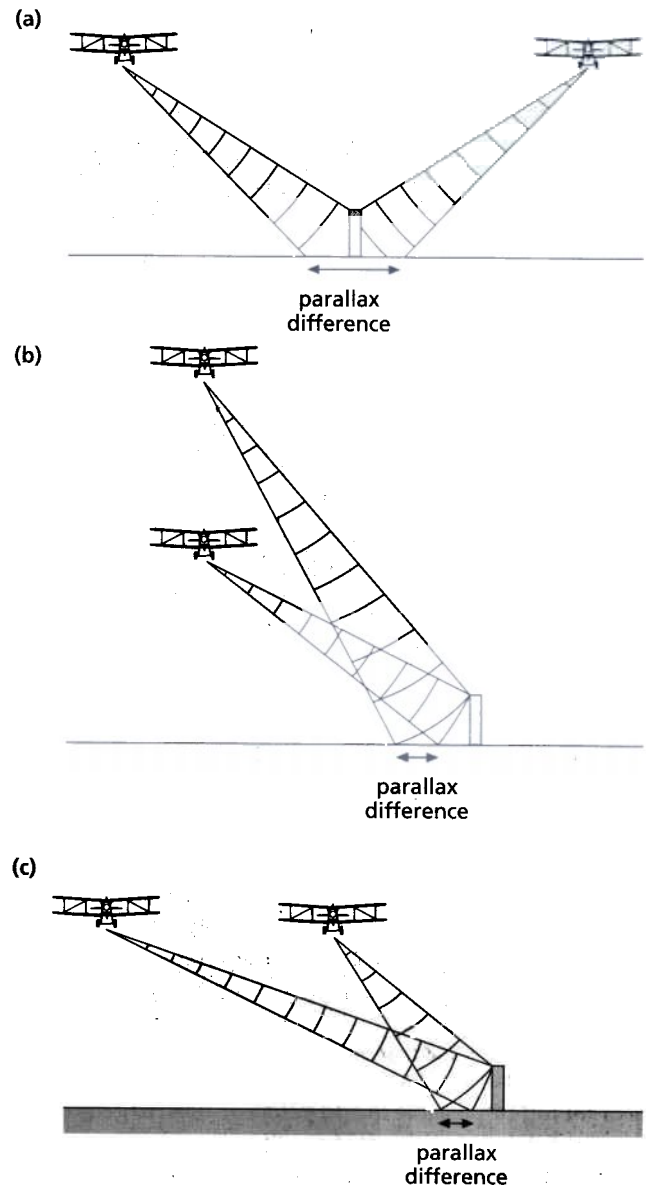


Fig. 7.21 Radar parallax suitable for stereoscopic viewing results from three main configurations: (a) illumination from opposite sides; (b) illumination from the same side but from different altitudes; and (c) illumination from the same side, at the same height but with different depression angles. Configurations (a) and (c) are possible from orbit. The method that produces best stereoptic models is (c).

relief. For this reason, the majority of stereoscopic radar missions adopt the same-side, same-height configuration (Fig. 7.21c). This is possible from a fixed orbit, provided the antenna can be pointed. Figures 7.22 and 7.23 give examples of stereoscopic radar images from aircraft.

7.2.5 Multipolarization radar

A variety of interactions between radar, the surface cover and subsurface materials can alter the polarization of

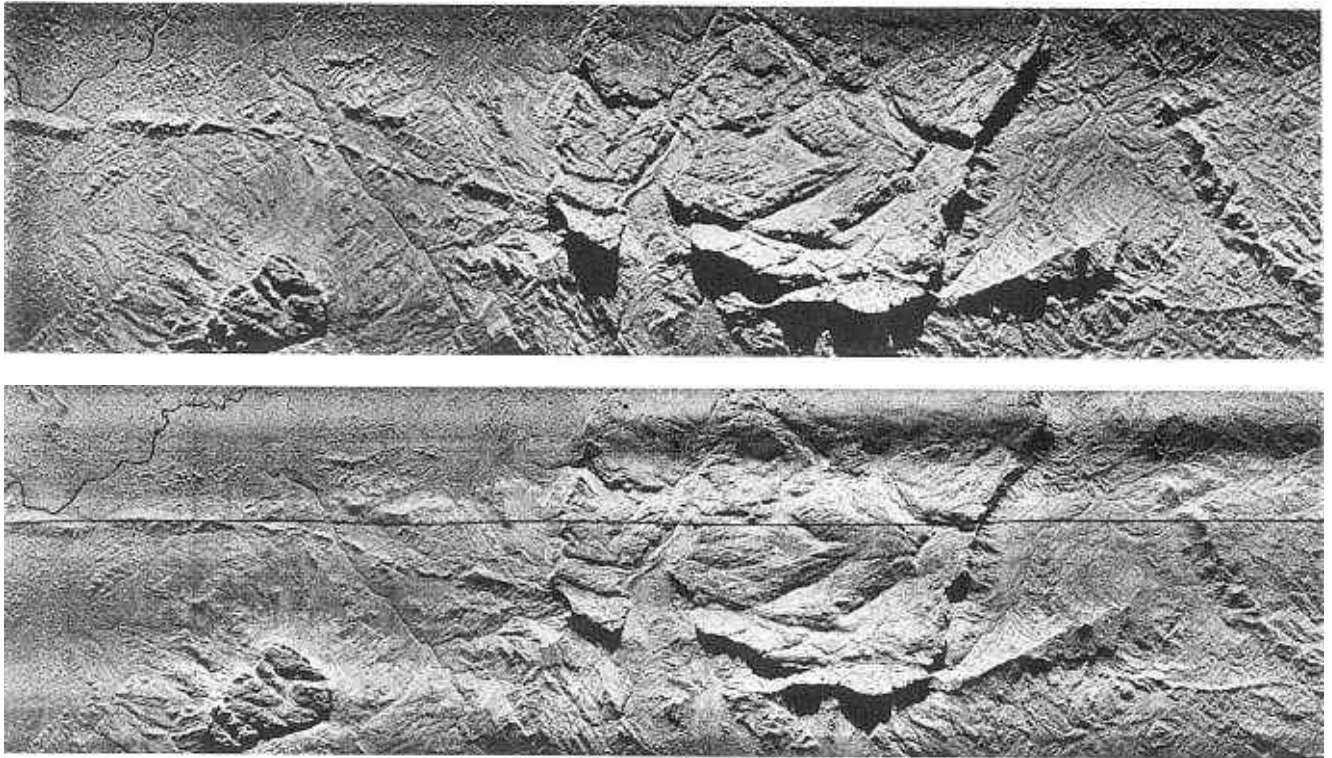


Fig. 7.22 In southern Venezuela, near to the Guyana border, the spectacular plateaux formed by fault-bounded blocks of the Cambrian Roraima Formation were the inspiration for Conan Doyle's 'Lost World'. Although not infested with dinosaurs, they are generally shrouded in cloud and comprehensive field work is not possible. This L-band stereopair, produced from an aircraft by the same-side,

same-height method, reveals many details of the local structure, including evidence for an unconformity where the Roraima Formation oversteps a dyke in the crystalline basement in the left-hand part of the pair. Courtesy of Earl Hajic, University of California, Santa Barbara. Images produced by Goodyear Aerospace Corporation.

radar waves (Section 7.1.3). By examining different combinations of transmitted and received polarizations it therefore should be possible to add yet more flesh to geological interpretation of radar images. In a way analogous to multifrequency radar it helps discern subtle differences in surface roughness, vegetation cover and subsurface granularity, which are not visible on single radar images or on those from other parts of the spectrum. Information of this kind is useful in an empirical sense for rock discrimination and highlighting anomalies. It is of more value to the botanist or agricultural specialist in quantitative measures of the kinds and densities of vegetation cover.

Different send–receive combinations can be compared as separate images, but as always, their combination as false-colour images makes far better use of the eye's acuity in comparative studies. Figure 7.24 and Plate 7.1 show examples of the potential of multipolarization radar images applied to the study of varied sequences of bedrock.

7.2.6 Multifrequency radar

The most powerful means of discriminating between surfaces with different degrees of roughness is to exploit

the wavelength dependency of the Rayleigh criterion (Section 7.1.2, Table 7.1). This requires illumination of an area by radar using several frequencies. As well as aiding the mapping of different lithologies and superficial deposits in terrain bare of vegetation, images of back-scattered energy for a range of frequencies can provide interesting information from densely vegetated areas such as jungles.

At relatively short wavelengths (Ka- and X-bands) the returns are mainly from any plant canopy that is present, rather than from the surface. The high dielectric constant of vegetation, owing to its water content, makes all parts of plants highly reflective to these wavelengths. Radar therefore is scattered primarily by leaves, and to a lesser extent by twigs and branches. At longer wavelengths (L-band) the dielectric constant is smaller and greater penetration is possible. This means that the back-scatter is influenced by the ground surface beneath grasses. In forests, even L-band radar fails to penetrate the canopy, and back-scatter is mainly from twigs and branches. Combining different wavelengths therefore should give information relating to three attributes of the surface—its topography, its roughness and the density of vegetation

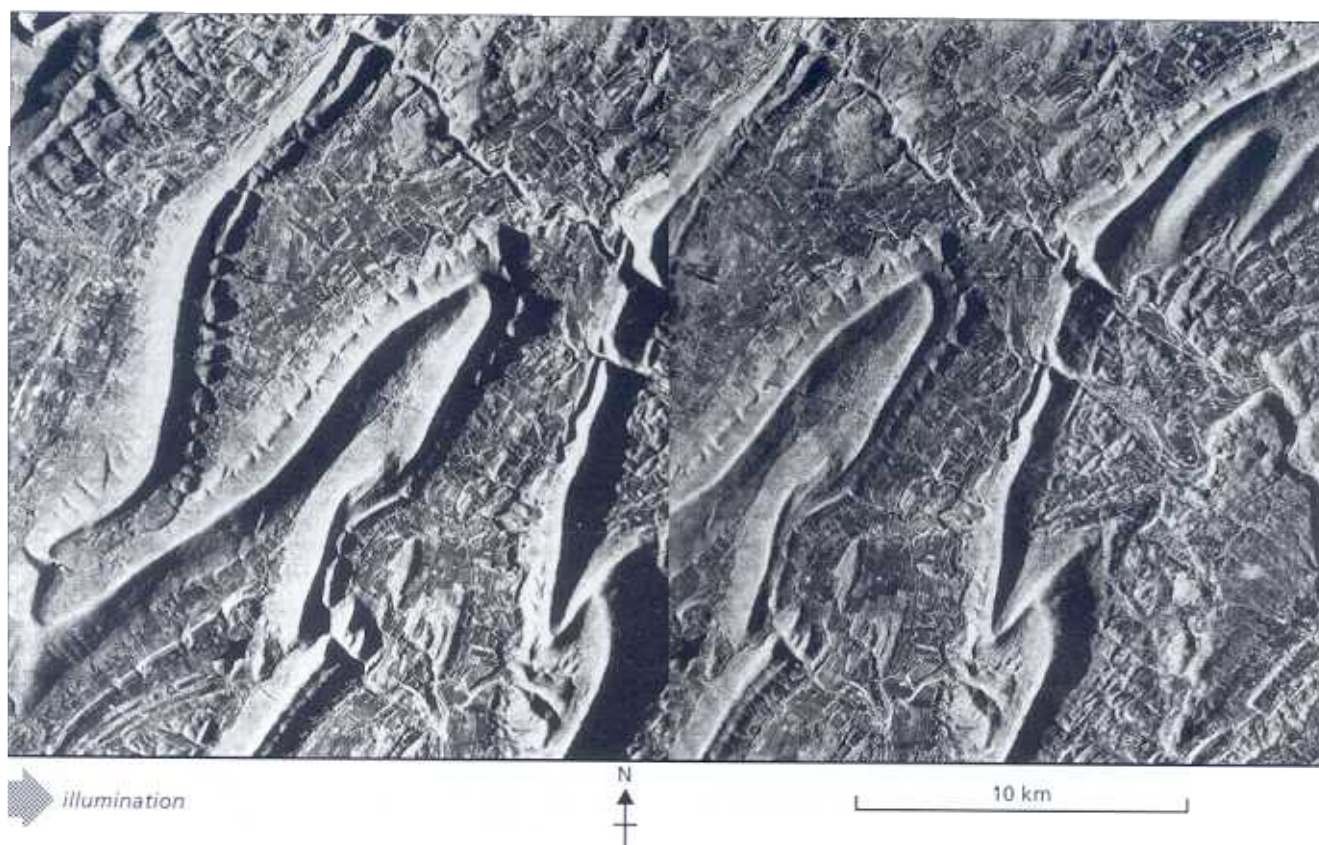


Fig. 7.23 This same-side, same-height stereopair of X-band airborne radar images is of part of the same area shown in Fig. 7.5. It shows considerable detail of the folding and faulting in the Carboniferous sediments of the Appalachians near

Pittsburgh, Pennsylvania, USA, and its quality is good enough for accurate photogrammetry. The images were produced by the STAR digital SAR system, and are provided by courtesy of Intera Technologies Ltd, Calgary.

cover. This potentially is of interest in the search for geobotanical features related to geochemical anomalies. It is certainly a means of looking at the different vegetation covers that sometimes distinguish one lithology from another.

Radar images of unvegetated areas remove the complicating effects of vegetation, so that varying responses to different radar frequencies are more easily interpreted geologically. Figure 7.25 shows three images that use different radar frequencies over a bare volcanic terrain.

Despite the restriction of its deployment to a limited number of orbits, the SIR-C/XSAR experiment carried by the Shuttle in 1994 produced the largest available archive of multifrequency radar images. Three wavelengths (L, 23 cm; C, 6 cm; X, 3 cm), together with various polarization modes provide a variety of RGB image possibilities. Because the Rayleigh criteria for each band (Equations 7.1 and 7.2) give > 7.5 , 1.9 and 0.97 cm for 'rough', and < 1.3 , 0.34 and 0.17 cm for 'smooth', respectively, for L-, C- and X-band wavelengths. The colour in such images is a sensitive guide to the nature of rock and soil surfaces. Plate 7.2 is a good example of how useful

such information is in providing an additional dimension for geological interpretations of an area.

7.2.7 Digital processing of radar images

Processing of digital radar data permits geometric correction and removal of noise (speckle) inherent to the way in which they are captured. Multifrequency and multipolarization data, combined as RGB images, lend themselves to various kinds of enhancement that helps make the images more easily interpreted visually. It might also seem that variation in DN, resulting partly from surface roughness, can contribute to classification when combined with other data. However, the dominant control over radar tone at all scales is topography. This has a much greater contribution than in images of reflected and emitted radiation. Consequently it tends seriously to disrupt the accuracy of classification in all except the most subdued terrains.

The main opportunity for automated information extraction from radar data lies in texture analysis and enhancement. There is far more textural information in a radar image than the eye can handle and sort

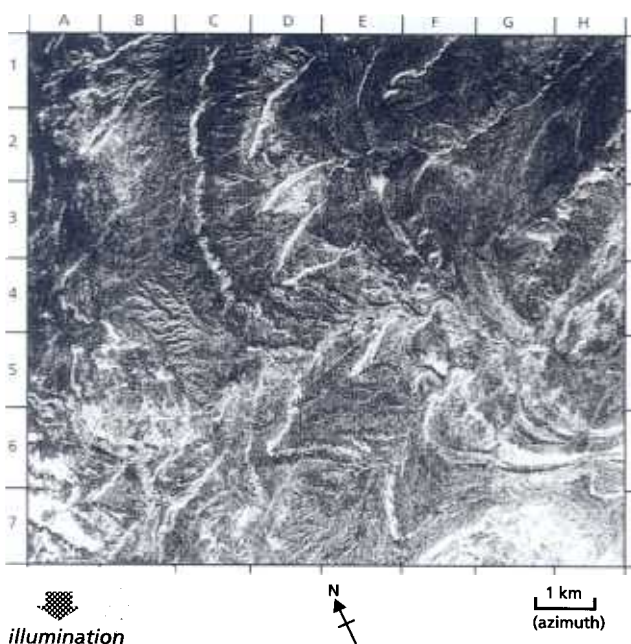
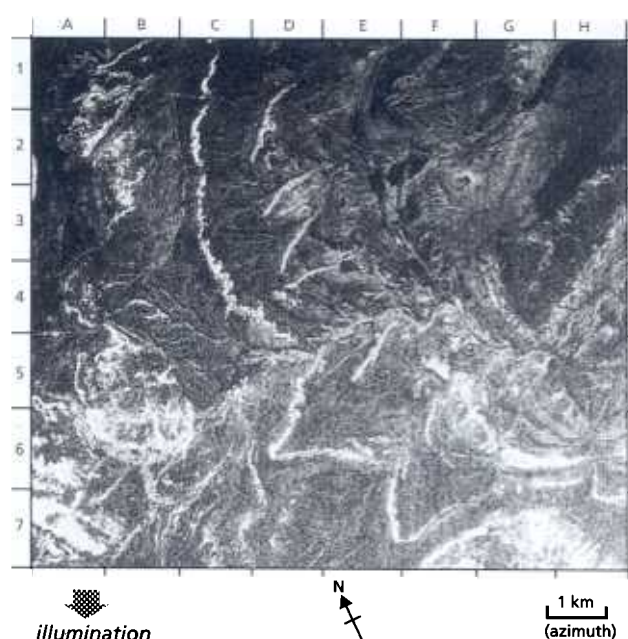
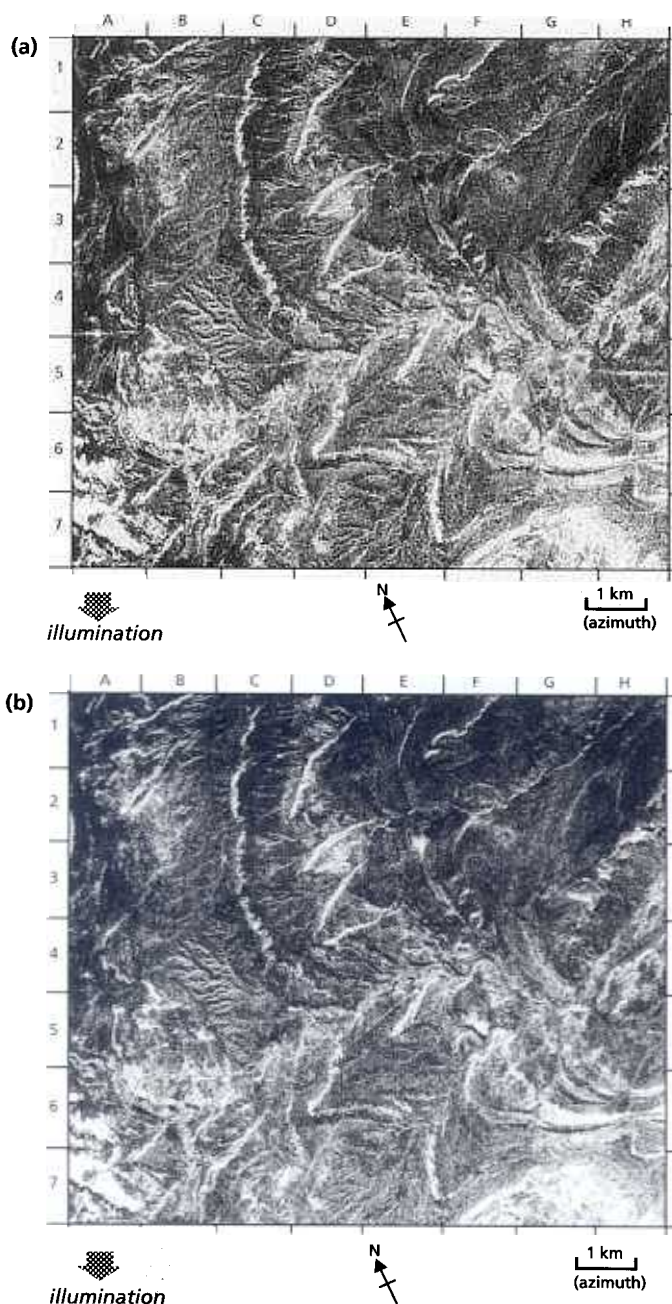
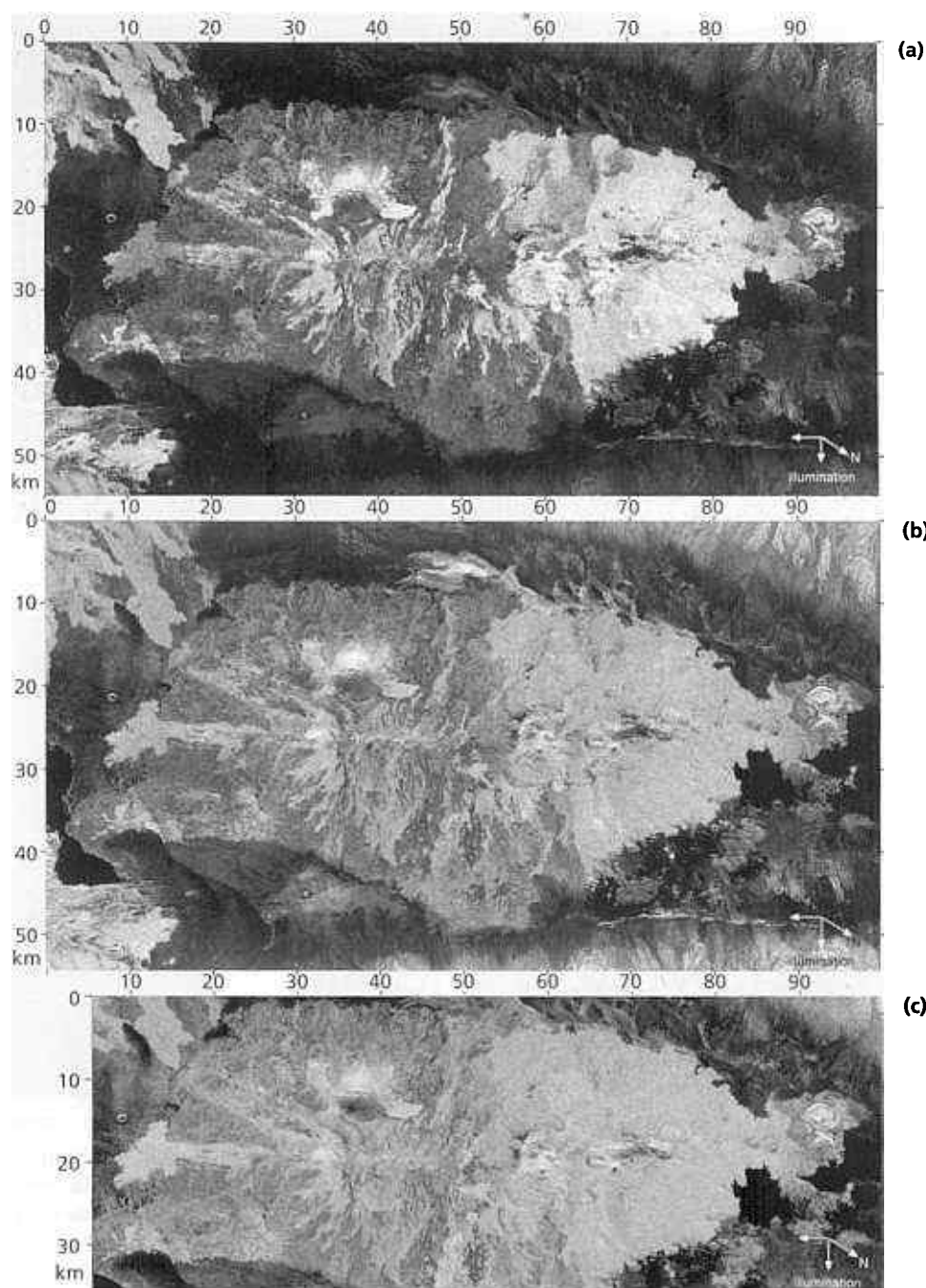


Fig. 7.24 These images are from an airborne L-band radar study of variations in back-scattered energy using various modes of polarization: (a) HH, (b) VV and (c) VH. The target area is in the Wind River Basin of central Wyoming, USA, and comprises a uniformly dipping sequence of Triassic to late Cretaceous limestones, siltstones, shales and sandstones. The oldest rocks outcrop in the west of the image, and the dip is roughly eastwards. The illumination direction was chosen to minimize the effects of topography, mainly north-south escarpments. At first sight there appears to be little difference between the images of different polarization arrangements. Indeed, in this form the data are highly correlated so that any one image is as useful as the others. However, combining the data in a false-colour image shows the subtle effects of different surfaces on polarization (Plate 7.1).

into categories. So, processing radar image textures by computer aims to simplify and categorize them. The most obvious treatment is to use edge enhancement and directional filtering (Section 5.3) to extract the maximum amount of information about curvilinear features that are present in an image. This, however, ignores a lot of the variation in the image that lies between linear features and is distributed in more or less homogeneous patches, and serves to decrease the already poor signal-to-noise ratio. To extract information relating to this aspect of the data means a more quantitative approach.

There are two main ways of quantifying spatial information, of which texture forms a major part. The simplest is spatial frequency filtering (Section 5.3). Tonal variations with low spatial frequencies are most usually influenced by the overall changes in roughness in an area. This is fundamental lithological and vegetation information. The higher frequency variations stem from all the different topographic features—scarps, ridges, valleys, gullies, pits and hummocks. They contain much of the structural and textural information. Separating the two gross components by low- and high-pass filtering (Fig. 7.26) enables a colour rendition to be extracted from a monochrome

Fig. 7.25 The area shown is the huge Erta' Ale volcano astride an active mantle plume in the Danakil Depression of Ethiopia. The images are from the 1994 SIR-C/XSAR mission of the Space Shuttle: (a) L-band (23 cm), (b) C-band (6 cm) and (c) X-band (3 cm), with flight and illumination direction, and north arrows at lower right of each. The marginal ticks have a 10-km spacing. The roughly elliptical volcano occupies the central part of all three images. It is surrounded by areas of alluvial sediments, and the black areas at the left and right of all three images are lakes. To some extent the brightness of surfaces on the volcano depends on the age of individual lava flows, as lava surfaces are often blocky and angular immediately after eruption, but degrade through weathering and erosion to smoother surfaces. The brightest surfaces are probably the youngest flows. Brightness of the volcano surface become greater and more uniform from L- to X-band, as a result of the decreasing dimension of the Rayleigh roughness criterion at shorter wavelengths. L-band is more effective at discriminating individual flows and their relative ages. Alluvial areas are generally smoother than bare rock, but the increase in overall brightness of the surroundings of the volcano from L- to X-band images shows how successively shorter wavelengths respond more to subtle variations in surface roughness. This is most noticeable at top right. The lack of any returns from the lakes indicates that the water surface was barely rippled at the time of the overpass. More information becomes apparent after combining the multifrequency data in a false-colour image (Plate 7.2).



image, often with spectacular results. The filtered data are expressed in the HSI system (Section 5.2.2), where different hues are interpreted more easily in terms of their control by terrain features (Plate 7.3). Separation into high-, medium- and low-frequency components is also possible, when the image can be reconstructed in colour through red, green and blue filters.

A more complex option is to express texture statistically. As texture amounts to the spatial variation of tone in an image, an appropriate measure is the variance of DN within an area. The higher the variance the more

widely and rapidly DN vary, and the finer and more contrasted the texture. Relatively smooth textures should be expressed by low variance. The procedure involves measuring the variance of DN in the surroundings of each pixel, and assigning that to the pixel as a measure of variance. The art lies in choosing boxes of the right size to express different textures properly. With the radar data in this form it is possible to display an image of variance. It is important to note that variance, like band ratios of shorter wavelength data, is little affected by variations in the angle of illumination.

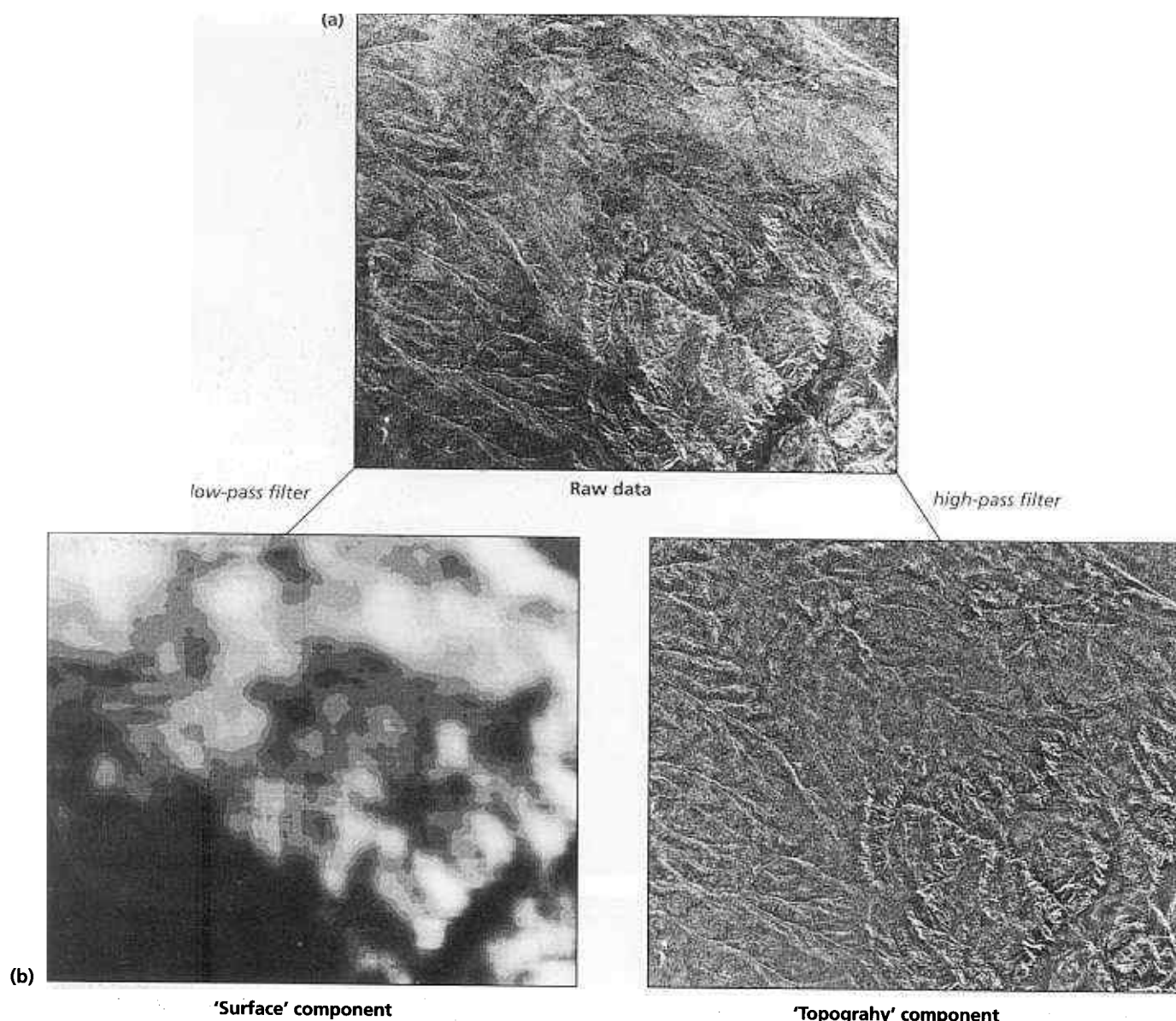


Fig. 7.26 Digital separation of different spatial frequency components from a Seasat image of the Wamsutter Arch in Wyoming, USA (a) results in the production of (b) an image of low-frequency variations related to gross variations in roughness and possibly rock type, and (c) an image of high-frequency variations containing topographic and textural

information. The high-pass image can be used to control intensity and the low-pass image to express hue in colour HSI image (Plate 7.3) where saturation is assigned a constant value to give the best colour rendition. Courtesy of Ronald Blom and Michael Daily, Jet Propulsion Laboratory, Pasadena.

The final step expresses differences in texture by several images that have been spatially filtered to enhance a range of spatial frequencies. Each is then recast as a variance image in its own right. The images are then treated as independent spectral bands, but as part of a spatial spectrum rather than an EM spectrum. These can be displayed together as colour images or used as a means of texture classification by multivariate methods (Section 5.6.3). Elegant as this approach might seem, little success has yet been achieved.

7.2.8 Radar combined with other data

Radar images are very different from those of other spectral regions, so it might seem appropriate to use them in the raw form together with other data sets as an uncorrelated dimension in classification. This is not a fruitful option. The great sensitivity to topography would mean that landforms played a greater role in classification than differences in surface properties. There would be gross misclassification. One option is to use measures

of topographic texture taken from radar images, such as the variance data described in Section 7.2.7. However, uses of this kind discard most of the visual information content inherent in radar images. An important use of radar images therefore is as a means of introducing or enhancing topography and texture into colourful, but less sharp combinations of other wavebands. The best way of doing this is through the HIS transform (Section 5.2.2), where radar data substitutes for the intensity in an RGB image of other data that contributes hue and saturation. The radar contributes a mass of structural information, the other data provide distinctions between different rocks (Plate 7.4). Chapter 9 shows other possibilities that introduce topographic information from radar into images of geophysical data.

7.3 Radar interferometry

You may be familiar with interferometry from optical physics. A puddle of water with a film of oil on it produces patterns of colour known as Newton's fringes on the surface. These bands result from light rays reflected by the smooth surfaces of the oil and the water beneath it. Tiny variations in thickness of the oil film cause interference between the phases of the two sets of reflected light, to give different colours. The patterns relate to variations in thickness of the film. Radar interferometry (sometimes given the acronym InSAR) is similar and generates interference patterns between two sets of radar signals scattered from the Earth's surface. The difference is that interference between the two signals results from the time lag between one signal and the other, both of which emanate from the same object on the ground. This results from the differences in length of the two radar paths. These lengths are measurable in terms of the radar wavelength. If they are exactly the same then the two waves will be in phase. If there is a difference, then the signals will be out of phase to some degree, and will interfere in the same manner as do light waves in Newton's fringes.

There are two common configurations for radar interferometry. Two antennas separated by an accurately known distance form the set-up for the fixed baseline method (Fig. 7.27a). An alternative is using back-scattered signals from overpasses on different dates.

Combining two interferometric radar signals produces an interferogram or fringe map, which looks similar to those seen in a film of oil. In the case of a fixed-baseline using two antennas the fringes represent differences in elevation of the surface, and can be recast as a digital elevation model (Chapter 8). Interferometric fringes produced by registering radar signals from two overpasses of a single receiving antenna, as on ERS-1 or -2,

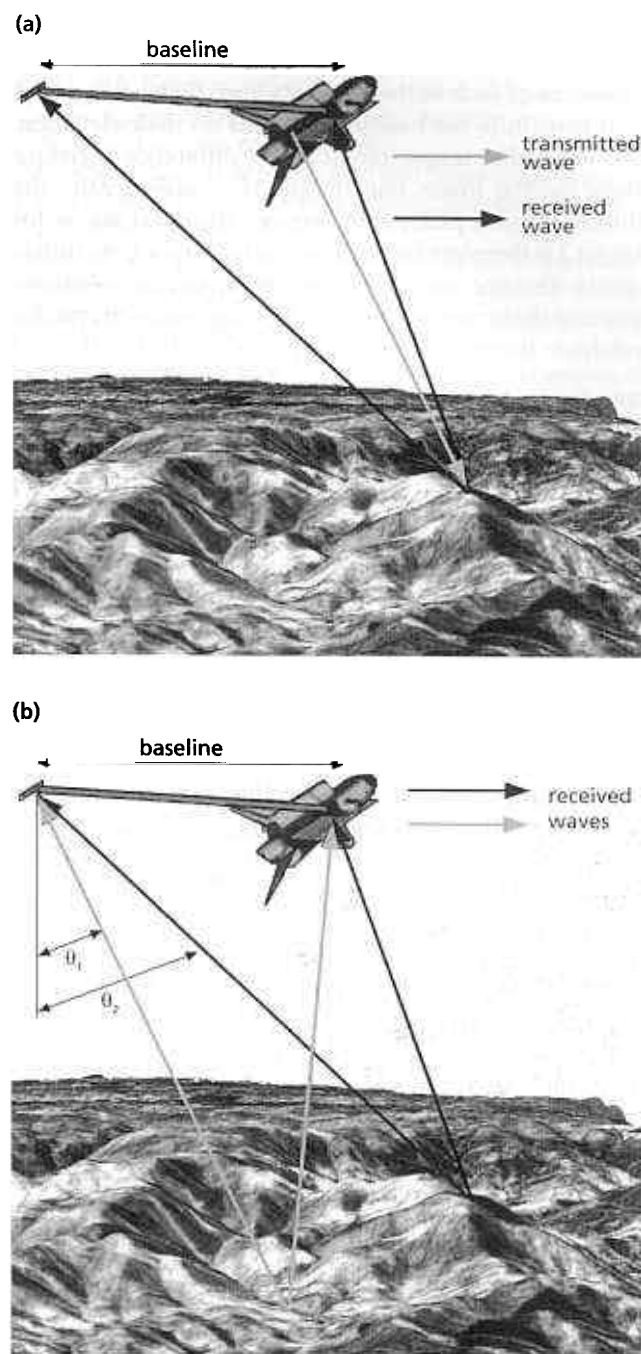


Fig. 7.27 (a) Antennas for fixed baseline radar interferometry (b) Radar wave paths to fixed-baseline antennas for points at different surface elevations.

represent any changes in elevation at different points on the scene. Such changes might occur during ground subsidence that results from landslips, mining or movement on a fault line, or because of inflation of a volcano as magma is emplaced prior to an eruption.

In order to see how fixed-baseline radar interferometry is sensitive to the height of the target, Fig. 7.27(b)

shows two different points with different elevations, in relation to radar waves. You can see that the different distances of each of these targets from the two antennas that constitute the baseline depends on their elevation. For the higher target (target 2), the difference is greater than for the lower one (target 1). Consequently, the interferometric phase shift between the radar waves for target 2 is therefore larger than it is for target 1. As differential distance gets larger so too does the incidence angle at the target (so, $\theta_1 < \theta_2$). The interferometric phase relates to the incidence angle by:

$$\text{interferometric phase} = B \sin(\theta) / \text{wavelength} \quad (7.3)$$

where B is the baseline length. Incidence angle (θ) can be calculated from the measured interferometric phase and known radar wavelength. The topographic elevation then stems from geometry. Accuracies of the order of a few metres are possible from orbit.

The calculation of changes in elevation from dual-date radar interferometry is more complicated. As the radar equivalent of Newton's fringes are in steps of one wavelength of phase shift, the changes can be assessed in millimetres—each fringe in a 10-cm radar system is divisible into many lesser steps.

The European Space Agency's ERS-1 is an operational source of dual- or multitime interferometry. NASA launched a Shuttle mission (the Shuttle Radar Topography Mission—SRTM) in February 2000, which carried a fixed-baseline InSAR system aimed at providing a digital elevation model for 80% of the Earth's land surface between 60°N and 56°S. The SRTM reused C- and X-band SAR instruments from the 1994 SIR-C/XSAR mission, with the addition of second antennas on a mast extending from the Space Shuttle's payload bay. The resulting DEMs will have a horizontal resolution of 30 m and one of up to 4 m for elevation. This comprehensive mapping to very high geodetic standards took a mission lasting only three weeks. By November 2000, NASA Jet Propulsion Laboratory and the German Space Agency DLR had released a few preliminary SRTM data, global coverage being scheduled for 2002.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file Menu.htm in the Imint folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

On the Menu page, select the link Miscellaneous: Image Data, where you will find examples of images based on a variety of SAR techniques. Of great interest are DEMs made using InSAR, which you can find in the link Miscellaneous: Digital elevation.

Further reading

- Amelung, F., Galloway, D.L., Bell, J.W., Zebker, H.A. & Lacznia, R.J. (1999) Sensing the ups and downs of Las Vegas: ISAR reveals structural control of land subsidence and aquifer-system deformation. *Geology* **27**, 483–486.
- Berlin, G.L., Schaber, G.G. & Horstman, K.C. (1980) Possible fault detection in Cottonball Basin, California: and application of radar remote sensing. *Remote Sensing of Environment* **10**, 33–42.
- Blom, R.G. & Daily, M. (1982) Radar image processing for rock-type discrimination. *IEEE Transactions, Geoscience and Remote Sensing* **GE-20**, 343–351.
- Blom, R.G., Crippen, R.E. & Elachi, C. (1984) Detection of subsurface features in Seasat radar images of Means Valley, Mojave Desert, California. *Geology* **12**, 346–349.
- Blom, R.G., Schenk, L.R. & Alley, R.E. (1987) What are the best radar wavelength, incidence angles and polarization for discrimination among lava flows and sedimentary rocks? A statistical approach. *IEEE Transactions, Geoscience Remote Sensing* **GE-25**, 208–212.
- Curlis, J.D., Frost, V.S. & Dellwig, L.F. (1986) Geological mapping potential of computer-enhanced images from the Shuttle Imaging Radar: Lisbon Valley anticline, Utah. *Photogrammetric Engineering and Remote Sensing* **52**, 525–532.
- Daily, M. (1983) Hue-saturation-intensity split-spectrum processing of Seasat radar imagery. *Photogrammetric Engineering and Remote Sensing* **49**, 349–355.
- Daily, M., Elachi, C., Farr, T. & Schaber, G. (1978) Discrimination of geological units in Death Valley using dual frequency and polarization imaging radar data. *Geophysical Research Letters* **5**, 889–892.
- Daily, M., Farr, T., Elachi, C. & Schaber, G. (1979) Geological interpretation from composited radar and Landsat imagery. *Photogrammetric Engineering and Remote Sensing* **45**, 1109–1116.
- Dellwig, L.F. (1969) An evaluation of multifrequency radar imagery of the Pisgah crater area, California. *Modern Geology* **1**, 65–73.
- Dellwig, L.F. & Moore, R.K. (1966) The geological value of simultaneously produced like- and cross-polarized radar imagery. *Journal of Geophysics Research* **71**, 3597–3601.
- Elachi, C. (1983) Microwave and infrared satellite remote sensors. In: *Manual of Remote Sensing* 2nd edn (ed. R.N. Colwell), pp. 571–650. American Society of Photogrammetry, Falls Church, Virginia.
- Elachi, C., Brown, W.E., Cimino, J.B., Dixon, T., Evans, D.L., Ford, J.P., Saunders, R.S., Breed, C., Masurski, H., Schaber, G., Dellwig, L., England, A., MacDonald, P., Martin-Kaye, P. & Sabins, F. (1982) Shuttle imaging radar experiment. *Science* **218**, 996–1003.
- Elachi, C., Roth, L.E. & Schaber, G.G. (1984) Spaceborne radar subsurface imaging in hyperarid regions. *IEEE Transactions, Geoscience Remote Sensing* **GE-22**, 382–387.
- Evans, D.L., Farr, T.G., Ford, J.P., Thompson, T.W. & Werner, C.L. (1985) Multipolarization radar images for geological mapping and vegetation discrimination. *IEEE Transactions, Geoscience Remote Sensing* **23**, 275–281.
- JPL (1980) *Radar Geology, an Assessment*. Publication 80–61, Jet Propulsion Laboratory, Pasadena, CA.
- JPL (1982) *The SIR-B Science Plan*. Publication 82–78, Jet Propulsion Laboratory, Pasadena, CA.

- Kasischke, E.S., Schuchman, A.R., Lysenga, R.D. & Meadows, A.G. (1983) Detection of bottom features on Seasat synthetic aperture radar imagery. *Photogrammetric Engineering and Remote Sensing* **49**, 1341–1353.
- Lowman, P.D., Harris, J., Masuoka, P.M., Singhroy, V.H. & Slaney, V.R. (1987) Shuttle imaging radar (SIR-B) investigations of the Canadian Shield: initial report. *IEEE Transactions, Geoscience Remote Sensing* **GE-25**, 55–66.
- MacKay, M.E. & Mougini-Mark, J.P. (1997) The effects of varying acquisition parameters on the interpretation of SIR-C radar data: the Virunga volcanic chain. *Remote Sensing of Environment* **59**, 321–336.
- Martin-Kaye, P.H.A. & Lawrence, G.M. (1983) The application of satellite imaging radars over land to the assessment, mapping and monitoring of resources. *Philosophical Transactions of the Royal Society of London Series A* **309**, 295–314.
- McCauley, J.F., Schaber, G.G., Breed, C.S., Grolier, M.J., Haynes, C.V., Issawi, B., Elachi, C. & Blom, R.G. 1982. Sub-surface valleys and geoarcheology of the Eastern Sahara revealed by Shuttle radar. *Science* **218**, 1004–1020.
- Pravdo, S.H., Huneycutt, B., Holt, B.M. & Held, D.N. (1982) *Seasat Synthetic-aperture Radar Data User's Manual*. Publication 82–90, Jet Propulsion Laboratory, Pasadena, CA.
- Rudant, J.P., Derooin, J.-P. & Polidori, L. (1994) Multi-resolution analysis of radar images and its application to lithological and structural mapping: Larzac (southern France). *International Journal of Remote Sensing* **15**, 2451–2468.
- Sabins, F.F. (1983) Geological interpretation of Space Shuttle radar images of Indonesia. *Bulletin, American Association of Petroleum Geologists* **67**, 2076–2099.
- Sabins, F.F., Blom, R. & Elachi, C. (1980) Seasat radar image of San Andreas Fault, California. *Bulletin, American Association of Petroleum Geologists* **64**, 612–628.
- Schaber, G.G., Berlin, G.L. & Brown, W.E., Jr (1976) Variations in surface roughness within Death Valley, California: geological evaluation of 25-cm wavelength radar images. *Geological Society of America Bulletin* **87**, 29–41.
- Schaber, G.G., Elachi, C. & Farr, T.G. (1980) Remote sensing of SP Mountain and SP Lava Flow in North-Central Arizona. *Remote Sensing of Environment* **9**, 149–170.
- Schaber, G.G., McCauley, J.F. & Breed, C.S. (1997) The use of multifrequency and polarimetric SIR-C/XSAR data in geological studies of Bir Safsaf, Egypt. *Remote Sensing of Environment* **59**, 337–363.
- Wadge, G. & Dixon, T.H. (1984) A geological interpretation of Seasat-SAR imagery of Jamaica. *Journal of Geology* **92**, 561–581.
- Wright, P. & Stow, R. (1999) Detecting mining subsidence from space. *International Journal of Remote Sensing* **20**, 1183–1188.

8 Non-Image Data and Geographical Information Systems

Earlier chapters have dealt almost exclusively with data that in one way or another were gathered in the form of images. These contain a wealth of information from which geological inferences can be made and synthesized in the form of maps. They are restricted to the use of EM radiation from the Earth's surface, however, and the spectral response of rocks poses fairly tight limits on geological interpretation. Except where structural features can be extrapolated in three dimensions, it also limits interpretation to, at most, a few kilometres of the crust, where uplift and erosion have produced extreme relief. Geologists are concerned with ever deeper levels, partly because the search for resources has to extend its scope as near-surface deposits become depleted, and partly because other kinds of data enable deep levels to be investigated. Indeed, the one presupposes the other.

Data that give clues to subsurface geology include well logs, seismic reflection sections, potential-field data from gravity and magnetic surveys, and results from various kinds of geophysical survey orientated at the electrical properties of rocks, soils and minerals. There are many other kinds of data with an indirect bearing on surface geology, such as measurements of emitted gamma-radiation, and geochemical surveys of rock and soil samples, vegetation, stream sediments and water. To these 'raw' data can be added information carried on existing maps. As well as defining lithological and stratigraphic units, and geological structures, maps can include sites of resource extraction, finds of economically interesting mineral associations (prospects or showings) and the range of information on surface elevation, hydrology, communications, habitations, land use and ownership generally associated with conventional topographic maps.

For decades geologists have used many of these other data sources in particular tasks, such as general mapping, geomorphological studies, exploration for hydrocarbons, metals and water, site investigations for engineering projects, and assessing environmental constraints on various operations, as well as in fundamental research (Chapter 9). The unifying factor among all these types of information is that they are geographical in nature. They relate to points, lines or areas on the Earth's surface expressed in various cartographic co-ordinate systems. As paper maps such data present a number of problems. They may be at a variety of scales and in different projections. Even as transparent overlays, the space required and the geologist's finite powers of mentally retaining information pose limits to integrating the information properly.

The symbolism used in maps varies (Section 8.1), and is sometimes out of tune with human visual perception. Many maps show variations in a third dimension as contour lines. Human vision treats lines and tonal variations in different ways (Chapter 2). Our perception of a contour map depends partly on the tonal variation produced by the spacing of contours—the closer the lines are the darker the appearance. The spacing of contour lines relates to gradients in the third dimension, however, so we see a contour map as an inverse representation of gradient. So even if contours are carefully labelled there is conflict between what is learned about contour maps and their interpretation and innate human perception. Contour maps are notoriously difficult to relate to the real world. Section 8.2 centres on converting information that is generally rendered as contour maps into image form. Images are more readily understood and amenable to enhancement using digital image processing.

The variety of data that a geologist may need to assemble, brood over and interpret can be represented as a series of layers (Fig. 8.1). Each layer contains information pertinent to a specific element of the task. To assess water resources for domestic and agricultural uses a hydrogeologist might need layers for geology, soil type, rainfall, land use, hydrology, data from wells, topography, land ownership, transportation routes and settlements. From these basic layers it would be necessary to extract information on the distribution of aquifers, well yields, stream gauging data, faults, topographic slope, the facing direction of slopes (slope aspect), and so on. This is a

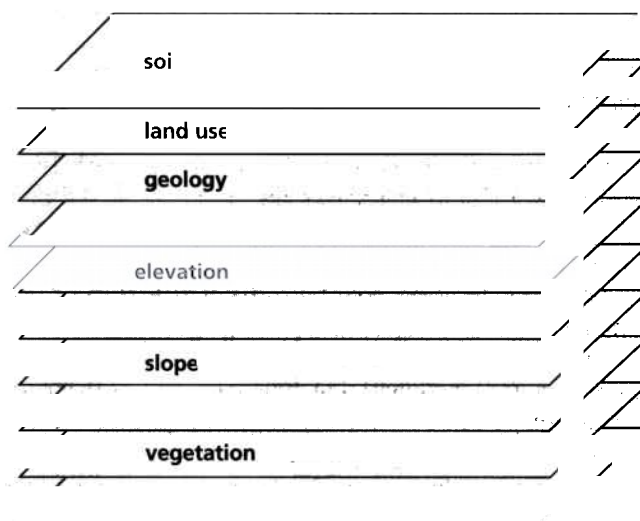


Fig. 8.1 Layered structure of primary geological and related data.

daunting body of conventionally separate information. An agriculturist working on the same programme would need some of this information too, plus more detail on, say, soil texture, soil depth, drainage properties, content of nutrients, pH, etc. Map information in paper form clearly poses problems for both, indeed, some information may well not appear on maps. Inevitably the analysis is slow and largely dependent on intuition and arbitrary selection of information.

An organized catalogue of geographical information as a series of exactly registered layers in digital form is the basis for management and interrogation of map data in a geographical information system (GIS). A GIS not only rationalizes data, but greatly speeds access to them. They become amenable to a variety of objective and efficient means of analysis, which encompass all pertinent information and thereby broaden the scope for interpretation, planning and execution. Section 8.3 introduces some of the basic concepts and techniques in GIS.

8.1 Forms of non-image data

The raw data available for use in geological and many other kinds of investigation can take a variety of forms:

- 1 values for some variable at a point, for instance elevation, geochemical and gravity data, borehole records, dip angle and direction;
- 2 continuous records of a variable along a line, as from seismic profiles, airborne and ground geophysical surveys;
- 3 areas designated as homogeneous categories, such as lithostratigraphic, soil and vegetation maps, and results of classification;
- 4 lines expressing relationships or linear categories—faults, intrusive contacts, roads and property boundaries.
- 5 continuous or semicontinuous grey-tone or colour images of remotely sensed data.

Types 1, 3 and 4 often involve information that is not part of some continuous variation, but expresses defined categories or attributes. Obvious examples are the metal-bearing minerals found at a mining prospect, names of rock types, different soil classes and various senses of fault movement. Moreover, a named category often has other associated attributes. A stream will have a particular order in relation to its connections with other larger and smaller streams—the higher the order the greater the number of tributaries—and will have a unique catchment size. It may be named in its own right or as a tributary of a larger system. A particular soil type may have variable colour, texture and drainage properties, and also can be categorized in terms of how prone to erosion it is. Roads have different numbers of lanes and various paving surfaces, as well as a designated code. Rocks have many attributes, a few being mineralogy, chemical composition, radiometric age, porosity and even associated plant community.

There are two basic models that encode geographical information, known as raster and vector models. A raster model represents the real world by dividing two-dimensional space into cells with definite shapes. Cells may be rectangular, triangular, hexagonal or any other shape that allows a systematic tiling or mosaic to cover the surface. Most rasters use square cells arranged in a grid of rows and columns, in the manner of a remotely sensed image. Their precision is limited by the size chosen for the cells. Objects and properties referring to them are represented by attributes assigned to the cells. A cell's attribute applies to every location within it, and the cell cannot be divided. The clearest examples of raster-format data are provided by digital images, including paper maps captured by various scanning devices. Maps are a great deal simpler than images, and comprise a limited number of divisions, compared with the 2^{24} divisions possible with a colour image of the Earth's surface. Raster models of maps might represent the boundaries between divisions accurately, but that is at the expense of redundancy for the areas that they separate; they are expensive in terms of file size. Other types of raster model are renditions of numerical nonimage data recast in the form of a grid of values (Section 8.2.1), which are more image-like than maps.

In the vector model, points and lines define the boundaries of objects or conditions, much as if they were drawn on a map. A small number of elements make up a vector model—points, lines and closed loops (polygons) made of straight-line segments. Lines are defined by their end-points or nodes, and by the vertices joining straight segments between the nodes, so that complex curves comprise many segments and vertices. Polygons are lines closed by a single node. The accuracy of vector-format data is limited only by the precision of surveying, the encoding length of computer words (now commonly up to 64 bits) and the number of points used to represent curved lines and polygons in the real world. Primarily, vector format represents spatial entities, but many attributes can be recorded for, and assigned to, each point, line or polygon within a single file. Data input to vector format can take many different forms, by keying in location and attributes at a terminal from field records, through data logging and positional devices such as a satellite positioning system used in the field, by manually digitizing existing maps, using complex pattern recognition to extract vectors from raster-format images, by line-following scanners, and by input of previously compiled vector files.

Figure 8.2 shows the essential differences between raster and vector formats. Raster data files, because they cover the whole of an area, each contain millions or tens of millions of cells. In this regard vector-format files are much more compact, comprising up to tens of thousands of elements. Each model has its advantages and

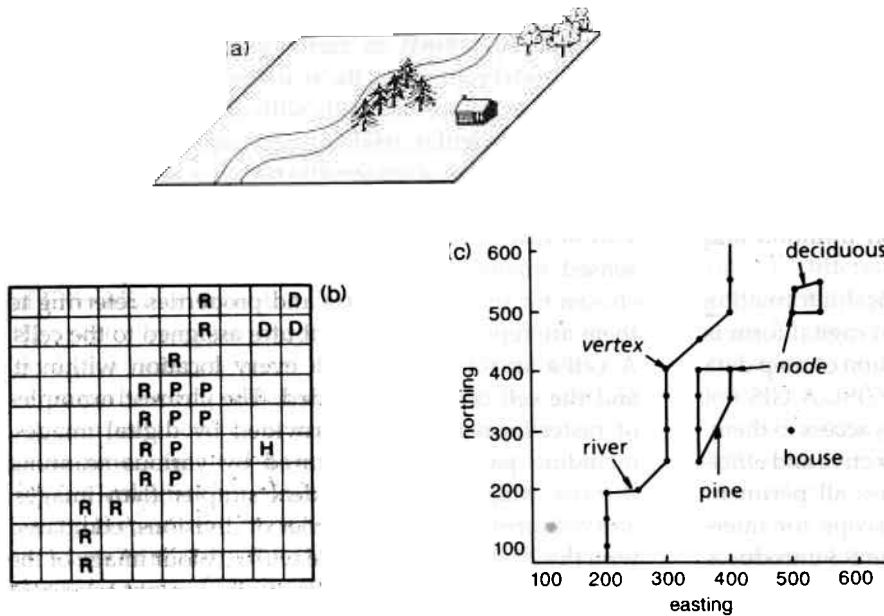


Fig. 8.2 Differences between raster- and vector-format data. A simple scene of reality (a) is represented as a raster by (b), where each element of the scene is made up of square pixels, and in vector format (c) by points, lines and polygons labelled with an attribute.

disadvantages. The raster model has a far simpler data structure than the vector model. It allows much easier combinations of different data than the vector model, represents spatial variability much more efficiently, and permits a great variety of digital manipulations and enhancements. Vector-format data, as well as being more compact, encode topological relationships whereas rasters do not. Being very precise representations of definite spatial features, vector files are better suited to graphical representation than rasters, which present boundaries with a blocky structure as a result of their being represented by arrays of cells. Both raster and vector data are used in geographical information systems (Section 8.3), but raster-format renditions of nonimage data have many uses in purely visual interpretation and combination with the information in remotely sensed images.

8.2 Non-image data in raster format

The simplest raster-format data are facsimiles of the red, green and blue components of a colour map, produced as raster files by a scanning device. Although in theory three-colour maps can be classified into categories comprising lithological units, boundaries and structural features, in practice the cartographic use of hachured tones, similar colours for different features and the presence of symbols, such as dip arrows, tend to thwart such efforts. A better representation of the information content in the map is achieved by using the vector model, but that entails manually digitizing boundaries if the map is available only as a paper print. The appeal of

raster maps lies in their retention of the cartographer's art and the ability to combine them with other kinds of raster data in composite images (Section 8.2.6).

All other data involving quantifiable, natural attributes, which vary continuously over an area, initially require conversion from records at points or along survey lines to the regular grid that is the hallmark of the raster model. This involves interpolation from the values at surveyed positions to areas for which there are no data.

8.2.1 Interpolation and gridding

Producing a raster rendition of nonimage data gathered at more or less irregularly spaced positions involves mathematically predicting from nearby measured values, and the overall variation, values that might be found within regularly spaced cells covering the whole area occupied by the data. This is interpolation. Predicting values that lie outside the surveyed area is extrapolation, which is rarely used, except in predicting values that are at the boundaries of the area of cover.

Interpolation relies on expressing the known data in the form of a three-dimensional mathematical model. The simplest approach is to assume that the value at any point is the same as that at the nearest surveyed station. This divides the area into polygons containing survey locations, the sides of which are equidistant between adjoining positions. The model consists of steps, and unless the original data were gathered in a rectangular grid, the polygons are more akin to a vector representation than a raster. This method is rarely useful in geological applications, and the favoured models embody predictions of gradual change.

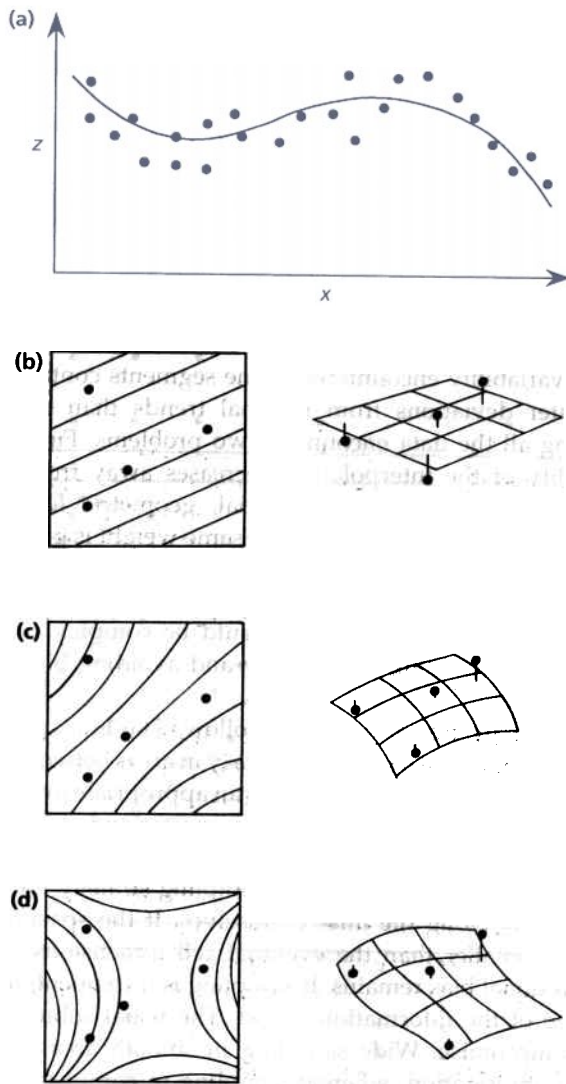


Fig. 8.3 (a) Best-fit line through points showing the variation of z with x . First-, second- and third-order trend surfaces fitted to four points are shown as contoured maps and three-dimensional surfaces in (b) (c) and (d), respectively.

The simplest of these predictive models is that involved in regression analysis, which fits surfaces to a data distribution to minimize the sums of squares of the distances between real points and the surface. Figure 8.3(a) shows a best-fit line that contains the variation of a property z with distance x . It is a cross-section of a simple curved surface in three dimensions. The distance in question is that between the points and the line in the z direction—there is no need to interpolate the distance, which implicitly is correct. The method uses polynomial surfaces of any order (Fig. 8.3b). This method, also known as trend-surface analysis, predicts the long-range variation of the third, nonspatial dimension and effectively smooths local variations, rarely passing exactly through the original data points (Fig. 8.4). Effectively, trend-surface analysis

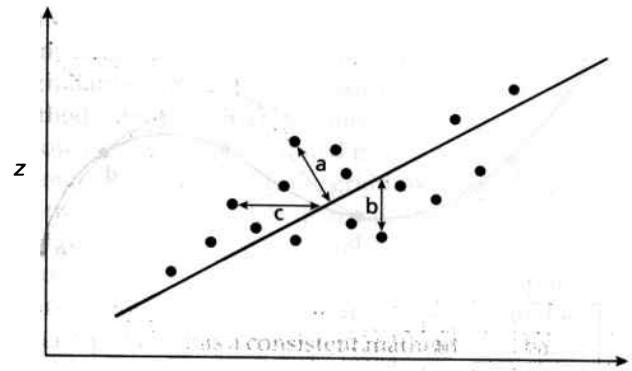


Fig. 8.4 Deviation of a best-fit line from the actual data points.

treats the 'raw' data as though they contain errors. Modern geophysical and geochemical techniques are remarkably precise, and the variability in data generally is real. So, a raster produced in this way degrades the information content. However, the deviations, or residuals, from trend surface models of different orders contain anomalous deviations from the general patterns represented by the surface itself (Section 8.2.5). A similar surface-fitting interpolation involves Fourier analysis (Section 5.3), but this too omits deviations from smooth surfaces.

Before digital fitting of curves to data, fitting was achieved using flexible rulers, technically known as splines. The important feature of a spline is that it can fit curves with a form that varies from place to place. Trend surfaces and Fourier series accommodate the whole in one complex function, and the points usually deviate from the computed prediction. A spline fit can be perfect, and the joins between one part of a curve and other different parts are continuous. This does not mean that a spline predicts exact values of the third dimension at unknown points, but adding extra points merely allows the spline to become a more precise reflection of reality. However, a spline makes no allowance for errors among the data. A line drawn using a spline is approximately made up of segments defined by different cubic polynomials, and has continuous gradients and gradients of gradient (first and second derivatives). Consequently, splines lend themselves to mathematical expressions, and can be applied to two- and three-dimensional data sets using a computer. For lines and surfaces these are known as cubic-spline and bicubic-spline functions, respectively. Mathematical splines can fit data exactly, or, if errors are suspected, can incorporate a degree of smoothing that allows for imprecision. Figure 8.5 shows a spline fit to several points, and the modification that must occur if one point is revised subsequently, emphasizing the limitations of this approach.

One of the problems with using bicubic splines in gridding, as implied by Fig. 8.5, is that they can introduce

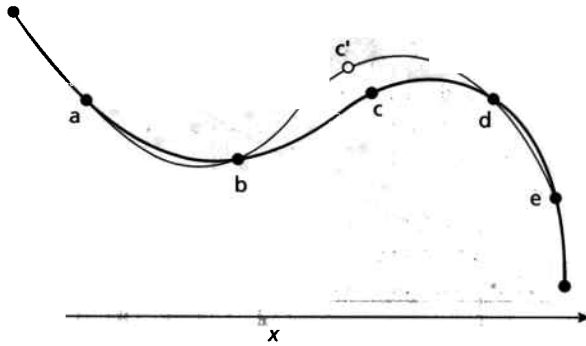


Fig. 8.5 Cubic-spline function fitted exactly to data points, the thinner line showing how the fit changes with the modification of one point (c to c').

artificial anomalies because of the way that they work. One means of avoiding this is to assign to an unknown cell the average value from actual points in its local neighbourhood. This is a moving-average approach, in which the size of the neighbourhood or 'window' can be varied. Another adjustment incorporates weighting the contribution of each point according to its distance from the unknown cell. However, this leaves several difficulties: What is the optimum size for the 'window'? What shape and orientation should it have? Is distance the best control over weighting? What errors are associated with the interpolated values? A mining engineer, D.G. Krige, addressed these problems in attempting to optimize predictions of spatial factors in mine planning. He recognized that many properties vary too irregularly to be modelled by smooth functions based solely on three-dimensional geometry. They require knowledge of some of the statistics involved. The mathematics involved is complex, but kriging, as this approach is known, allows great control on the precision and accuracy of gridding, as is needed in mapping variations of metal grade in ore bodies and in the production of digital elevation data (Section 8.2.4). However, methods based on splines and weighted moving averages tend to be the most frequently used means of producing rasters.

8.2.2 Geophysical data

Today, virtually every geophysical survey records data in digital form, with extremely precise geographical positioning by satellite navigation systems. The raw data therefore are available immediately—given payment of the appropriate fee—for interpolation using the methods outlined in the last section. Gridding may already have been done and the results stored, because this is essential in the computer production of conventional contour maps from survey data. There are, however, a few remaining problems. As point measurements generally are produced

with a fairly even spacing, problems are minimal with gravity data and similar information measured at discrete points. Many surveys are conducted along survey lines, however, either continuously, as with aeromagnetic surveys, or at regular along-line sampling intervals, as in airborne gamma-ray spectrometry and electrical ground surveys.

The main problem with line surveys is that the sampling interval along lines generally is spaced far more closely than the separation between lines—there is a distinct linear bias to the data. Moreover, in terms of the variability encountered, some segments contain far greater deviations from regional trends than others. Using all the data encounters two problems. First, the quality of the interpolation decreases away from the lines, and that imparts artificial, geometric features parallel to the lines. Second, the same weight is given to 'smooth' areas as to complex areas, thereby involving much redundant information and tending to extend smoothness into areas that should be complex. Some means of reducing redundancy and avoiding bias has to be sought.

There are two strategies to follow. The line spacing limits the degree of spatial accuracy in areas between the lines, and so the operator selects an appropriate grid-cell size. In general, a size between a quarter and an eighth of the line spacing avoids artefacts parallel to the lines. Another option is to adopt a sampling strategy for the raw data along the lines themselves. If the spacing is much smaller than the eventual cell dimensions then directional bias remains. If sampling is infrequent, then some of the information is lost. The usual solution is a compromise. Wide sampling in smooth areas loses little information, whereas sampling in complex areas concentrates on local minima and maxima, and, more important, on defining the shape of the variation where there are changes in gradient. Doing this automatically involves incorporating measures of local complexity, such as variance and the second derivative, plus identification of maxima and minima and an upper limit imposed on sampling density. Inevitably, some information is always lost, to be replaced by interpolated values, but this is offset by the advantages of rendering the data in raster form.

Much geophysical information dates from before the 'digital revolution'. Raw data as analogue profiles along survey lines needs conversion to digital form, either manually or using line-following software. Contour maps may well have been produced by manual interpolation. As well as the visual disadvantages of contour maps discussed earlier, their very nature in expressing continuous variations in discrete steps degrades the original data. In such cases the only option is to extract information by digitizing contoured data. Figure 8.6 shows three methods that differ in sampling strategy.

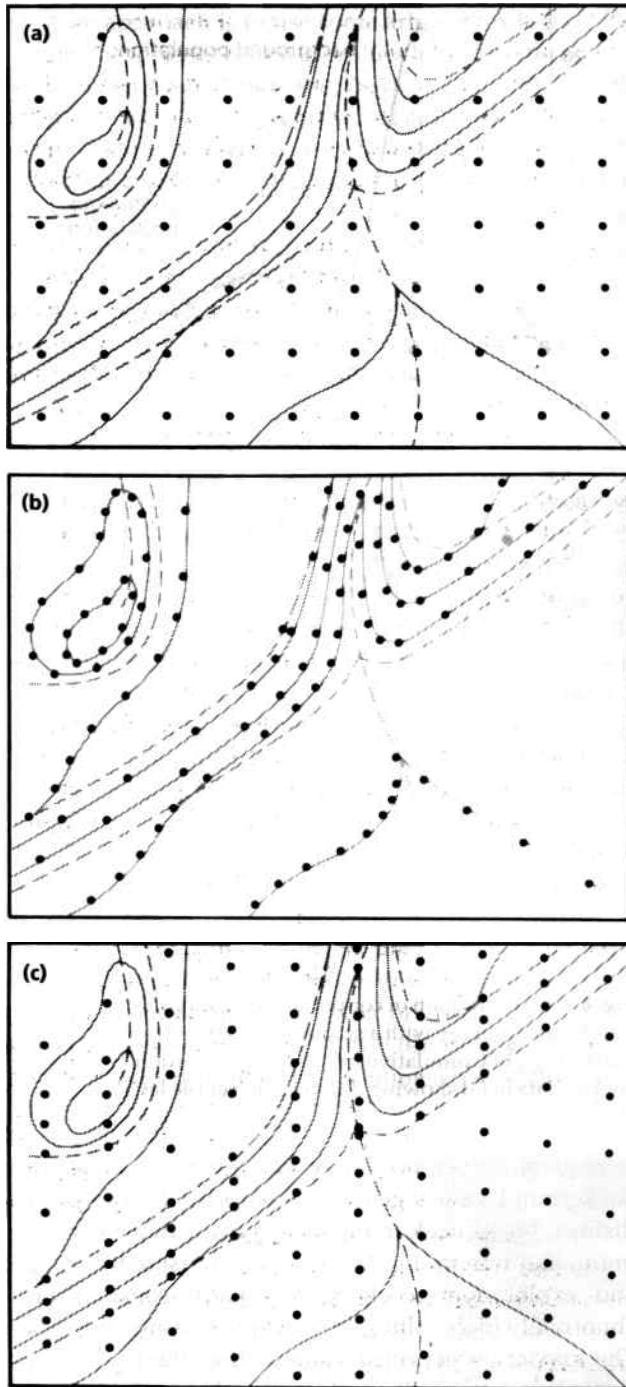


Fig. 8.6 Three methods of digitizing data from contour maps.

The first systematically estimates a value for the variable at points in a regular grid (Fig. 8.6a). Using the reasoning that more points are needed to define areas of increasing complexity means sampling with variable spacing (Fig. 8.6c). Both methods 'second guess' the methodological approach used in compiling the original map, inevitably assuming linear variation between contours in manual digitizing. The third method assumes that the

contours are positioned accurately, and samples along them (Fig. 8.6b). This maintains, and even extends the degradation implicit in contouring. However, the last method can be partly automated by applying line-following software to scanned maps, and the variations in contour spacing assists the gridding software in fitting a better surface to the data than is possible using the other methods. None have particular advantages, but because producing raster maps generally involves input from many contour maps, the same method is essential so that the end product has a consistent mathematical basis.

8.2.3 Geochemical data

Measurements of the concentration of an element in a sample of rock, soil or vegetation relate only to the geographical location of the sampling. Gridding treats them in exactly the same way as geophysical data from point surveys. However, such surveys are slow and expensive to conduct and rarely figure in regional evaluation. A much quicker, cheaper and convenient approach exploits the natural sampling of soils and rocks by erosion, transport and deposition by flowing water. Most of the continental surface has drainage channels that are easy to locate on maps or aerial photographs, and easy to sample. Results of drainage geochemical surveys, using analyses of either stream sediments or stream water, are by far the most widespread kind of geochemical exploration. However, they present a peculiar problem in conversion to raster format.

The concentration of a chemical element in stream sediment or water does not represent the composition of the soil or rock in the immediate vicinity of the sample point. Instead, it represents the entire area supplying water and sediment moving past the sample point. That is, it represents the catchment area upstream of the sample point. Figure 8.7 illustrates the difficulty. The area represented by each sample increases the further down the drainage the sample is collected. This relates to the order of the stream—the more tributaries enter a stream, the higher is its order and the larger its catchment area. Consequently, in sampling all streams, irrespective of stream order, each sample not only has its own unique concentration but a unique significance in terms of the area that it represents. It is impossible to interpolate such results to a regular raster that has any useful meaning. In fact it would be grossly misleading to do this. The only valid representation would be assigning the measured values to the associated catchments. If there are several sample points along a drainage this clearly leads to great confusion. Furthermore, as the order of a stream increases, so the chances of anomalous values decrease owing to the diluting effect of water and sediment being contributed from a wider area. Much metal exploration involves drainage geochemical surveys

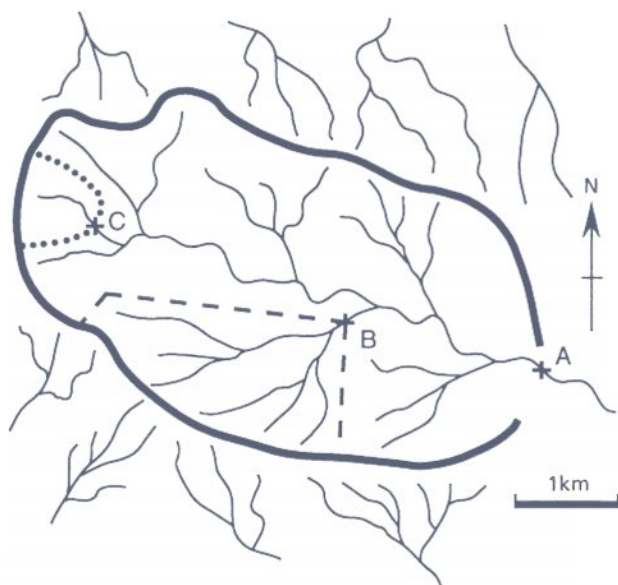


Fig. 8.7 Map showing the variation in size of catchments (dotted, pecked and bold lines) for points on streams of different order.

of this style, seeking anomalies low in a drainage and then tracing them up tributaries to their source. That largely rules out the data being useful in raster form. However, there is a strategy to overcome the problem.

The lower a stream's order, the smaller is its catchment and the more likely are the concentrations of chemical elements in stream samples to represent nearby rocks and soils. Provided sampling involves only streams of the same order, preferably the first or second order, and interpolation uses grid cells that have a similar size to that of the associated catchments, the resulting raster is realistic. Because of the wealth of possibilities for display and enhancement of geochemical data, and their combination with other kinds of data, either in raster images (see Plate 8.3) or GIS, this form of sampling is becoming the norm for drainage geochemical surveys.

Multi-element geochemistry is a powerful means of discriminating rock types and gaining insights into petrogenetic relationships, because of the different behaviours of various suites of elements during chemical fractionation by different geological processes. So, one use of raster-format geochemical data is in basic geological mapping, when the full range of variability is used. Geochemical exploration, primarily for metals but sometimes for hydrocarbons, relies on locating areas where concentrations of some elements rise well above the general levels or backgrounds associated with common rock types to form anomalies. A problem lies in deciding which values constitute anomalies within the full range of variation. A variety of solutions centre on statistical analysis of the data. The most frequently occurring ranges of values constitute the background, whereas

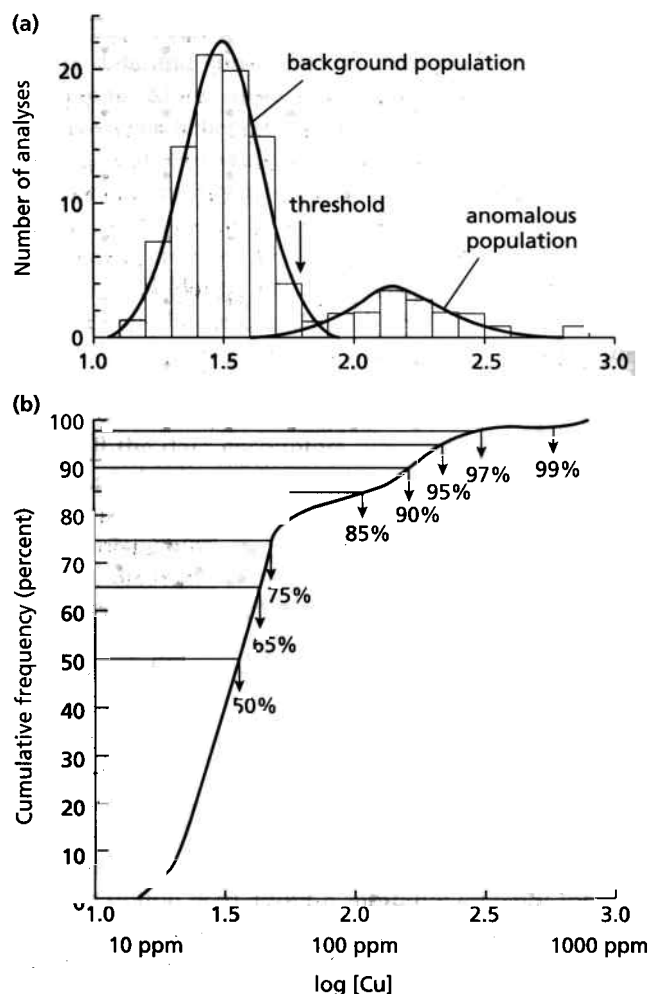


Fig. 8.8 (a) Histogram of copper concentrations in a geochemical survey with a smooth curve fitted to the variability. (b) Cumulative frequency distribution for the copper data in (a) showing different percentile levels.

infrequent occurrences may be significant anomalies. Background values generally comprise a more or less distinct, broad peak in the histogram with a clear maximum and minimum. Anomalies lie outside this range, and exploration geochemists usually consider only abnormally high values, for obvious reasons (Fig. 8.8a). The maximum expected value within the background range is sometimes termed the threshold, and it is possible to separate anomalies from the normal by setting all cells with values below the threshold to an arbitrary level and using contrast enhancement to highlight anomalies. More sophisticated statistics help refine this screening. One is to use the cumulative histogram and density slice at a variety of percentile levels (Fig. 8.8b). As many elements show a log-normal distribution of values—many occurrences over a restricted range (such as 10–50 parts per million) and a few occurrences over a very wide range (such as 100–1000 parts per million)—the histogram is often used with a logarithmic axis.

Another approach is to use some arbitrary rule that, for example, expresses cells with values below the threshold (t), between t and t plus the average value (a), $t + a$ to $t + 2a$, $t + 2a$ to $t + 3a$, and greater than $t + 3a$ with different intensities or colours, each constituting anomalous cells of increasing potential significance. Because of the wide variation in chemical composition of common rocks, particularly for some trace elements, a concentration that is anomalous for one rock type may be within normal bounds for another. So, a further complicating factor in using geochemical data is screening them according to the known distribution of different lithological units.

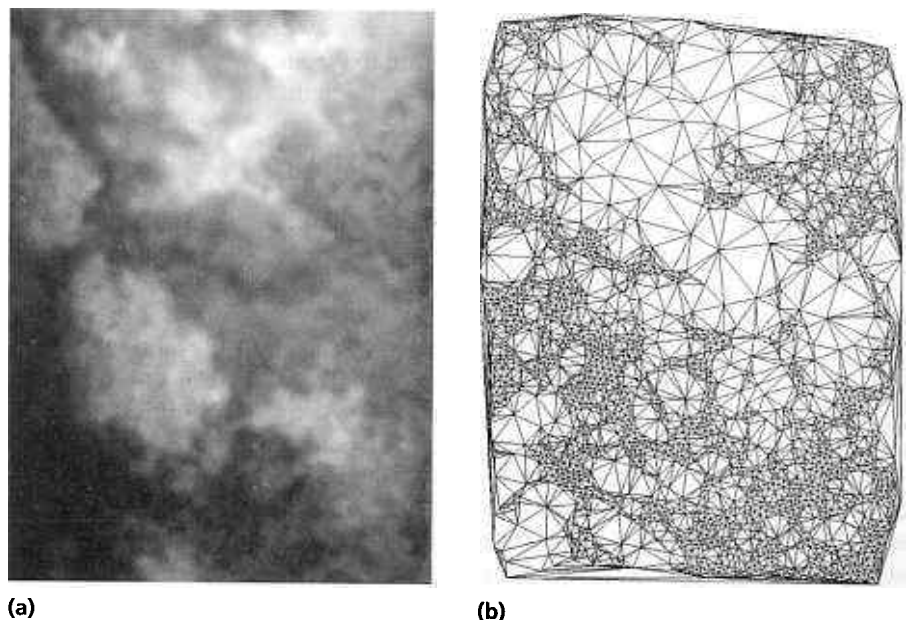
8.2.4 Elevation data

The Earth's surface is the most familiar example of continuous variation in three spatial dimensions, and is almost universally represented by contoured maps of topographic elevation. As well as containing accurate representations of topography, in which there may be clues to geological features that exert a control over land-forms, topographic data have immense potential for the extraction of other information with a variety of uses. Topography controls the rates and directions in which surface water flows, and it involves drainage networks that can be established without hydrological survey. Where there are known geological boundaries, which perhaps appear on remotely sensed images, their trends in relation to topography define their dip and strike, and even structure contours from the simple three-point solution. Topographic slope estimates give measures of the likelihood of soil erosion or risk from landslip. Incorporating topographic data with other kinds opens up new ways of visualizing relationships, such as per-

spective views of unknown terrain, introduction of relief displacement into remotely sensed images to give stereoptic viewing potential and the correction of remotely sensed data for variable solar illumination.

Digital elevation models (DEM) take a variety of forms. The simplest are raster arrays of elevation values, but these use the same density of cells irrespective of the complexity of the topographic surface, and so occupy large storage volumes. A number of methods use variable densities of data points to reduce storage space, but all must be gridded if the data are needed in raster format. The most commonly used of these efficient storage formats is the triangular irregular network (TIN). A TIN comprises a number of triangular facets (technically known as Delaunay triangles) linking points with known elevation, the greater the relief, the larger the number of points and facets. Usually, producing a TIN involves converting a DEM to the low density format (see Fig. 8.9). Capturing elevation data from existing contour maps and interpolating them can use any of the methods described in Section 8.2.1 and 8.2.2. For some areas DEMs in various forms are available from state mapping agencies. The US Department of Defense's National Intelligence Mapping Agency (NIMA) maintains a global archive of DEMs derived from surveillance satellites and aircraft for most of the planet at very high resolution. This is known as Digital Terrain Elevation Data (DTED) Level-1C. At the time of writing DTED Level-1C for areas outside the USA was accessible only to US Government Agencies or those of allied powers, but this may change if pressure from potential users is applied strenuously and continuously. A Shuttle Mission in February 2000 deployed an interferometric radar system to measure topographic elevation with a geographical

Fig. 8.9 (a) Digital elevation model shown as a grey-tone image. (b) Triangular irregular network (TIN) representing the same surface. Note how the density and size of triangles in the TIN reflect varying degrees of complexity in the digital elevation model (DEM).



resolution of 25 m, and a vertical resolution of 4 m. Early results from the Shuttle Radar Topography Mission (SRTM) show that the quality of the DEMs that it will produce surpass those previously restricted by NIMA.

Most topographic maps use elevation data derived from stereoscopic measurements of aerial photographs, using a variety of manual and semiautomatic photogrammetric instruments, backed up by precise geodetic surveying. The methods rely on measurement of the parallax differences inherent in the relief displacement in stereoscopic image pairs (Appendix A). Satellite systems, such as SPOT and JERS-1, produce digital stereopairs and it is possible to digitize photographic images accurately. This makes it possible to derive DEMs automatically, through the use of complex software. The parallax estimation that is essential in such automated cartography relies on the computer being able to recognize the same features on each image of the stereopair. An operator identifies a number of points on the stereopair that act as 'seeds' for pattern recognition, which builds up many more such ties points automatically. The parallax differences associated with these points form the basis for calculating elevation relative to accurately surveyed ground-control points. As elevation precision depends on that associated with parallax, the spatial resolution of the imagery is crucial, because parallax can be measured only in multiples of pixels. For 10-m resolution SPOT images, elevation precision is around 20–30 m—too poor for all but qualitative use of the resulting DEMs. Aerial photographs have resolutions in the 0.1–1 m range, and so provide DEMs that are suited to detailed cartography.

In the absence of readily available, precise DEMs, the best means of producing them for detailed, quantitative applications is by painstaking digitization of contours on large-scale ($> 1 : 25\,000$) maps and then interpolation (Section 8.2.1). This should ensure vertical precision of the order of a few metres.

Because the travel time of radar waves scattered from the surface can be measured very precisely, it might seem reasonable to expect that radar could be used for measurements of surface elevation to within a few centimetres. Seasat and ERS-1 deployed such radar altimeters, but results are available only for the ocean surface. The problem over land lies in both the difficulty in focusing the radar beam and in the response from rugged surfaces. From orbital altitudes the radar beam spreads to several kilometres, thereby limiting the resolution in the horizontal plane. Within this 'footprint' the highest features reflect the radar wave first and the lowest give a later return of the radar pulse. Consequently, all that can be measured in the vertical dimension is the average elevation and the range of elevations. For such a wide 'footprint' that information is of little use over all but the most gentle terrain. Radar altimetry therefore is of little use over land, waves at the ocean surface are

rarely higher than 1–20 m, however, and the average ocean-surface elevation can be measured to within a few centimetres within the wide radar 'footprint'. Wave height obviously is an interesting measurement for oceanographers and meteorologists, but the usefulness of ocean-surface elevation to geologists is not immediately obvious.

Sea level is closely controlled by gravitational factors, hence the diurnal and monthly patterns of tides. But the Earth's gravitational field also plays a role, as well as those of the Moon and Sun. Averaged over several months the ocean-surface elevation mainly relates to variations in terrestrial gravity. Over the oceans the gravitational field, and therefore the surface elevation is dominated by variations in ocean depth. Because water is about a one-third as dense as the rocks of the oceanic lithosphere, the shallower the water the higher the gravitational potential and vice versa. This means that water flows to shallow areas to form a bulge, and conversely the water surface is depressed over deep ocean basins. A feature on the ocean floor with a relief of 1 km results in a 2-m disturbance of the surface. With a vertical precision of 5 cm, radar altimetry therefore can represent ocean-floor features as small as 25 m, provided that they extend over several kilometres. So, as a result of this convoluted relationship, images of radar altimetry data over the oceans form bathymetric maps of such precision that subtle features of the ocean floor are readily apparent (Fig. 8.10). Because conventional bathymetric maps are compiled from soundings taken by ships, their quality varies according to the density of marine traffic. Over most of the oceans bathymetry derived from radar altimetry is far more informative. It is possible to use conventional bathymetry to remove the effect of varying water depth to analyse changes in gravitational potential resulting from variations in the density of the underlying crust and mantle. This can be used to map features possibly related to thermal convection in the mantle, and to locate deep sedimentary basins in areas of continental shelf.

Although radar altimetry appears unlikely to generate useful DEMs over land, because of beam-focusing problems, laser altimeters potentially can produce very high resolution in all three dimensions over land. Indeed, laser methods are used from orbit for precise site positioning in studies of plate dynamics. There is one major difficulty of a nonscientific nature associated with very precise DEMs. They are essential for the guidance systems of cruise missiles, as well as other military applications, and therefore are shrouded in secrecy. As cruise missiles have been around for well over a decade, so too have DEMs with the requisite precision.

With SRTM interferometric radar data (Section 7.3), all this set to change rapidly, provided the results are not placed under a security blanket when their quality and information content become clear.

Fig. 8.10 Image of sea-surface elevation derived from Seasat radar altimetry data, showing details of ocean-floor topography as a result of the effect of gravitational potential on the sea surface.

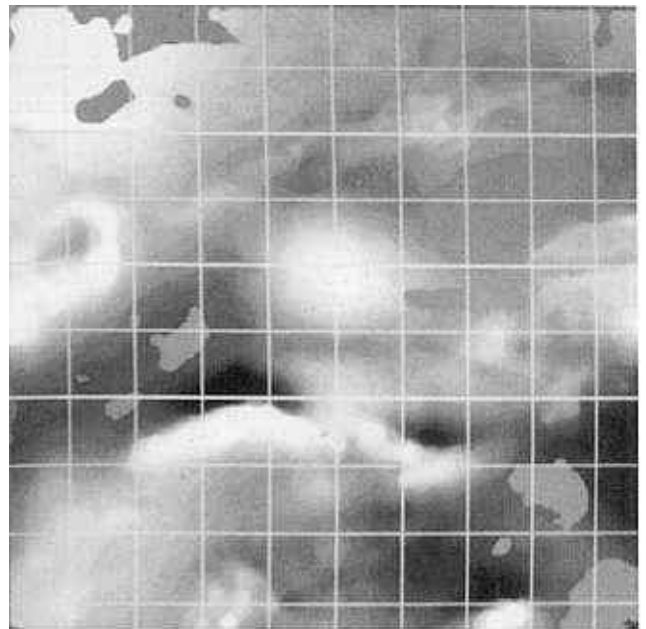
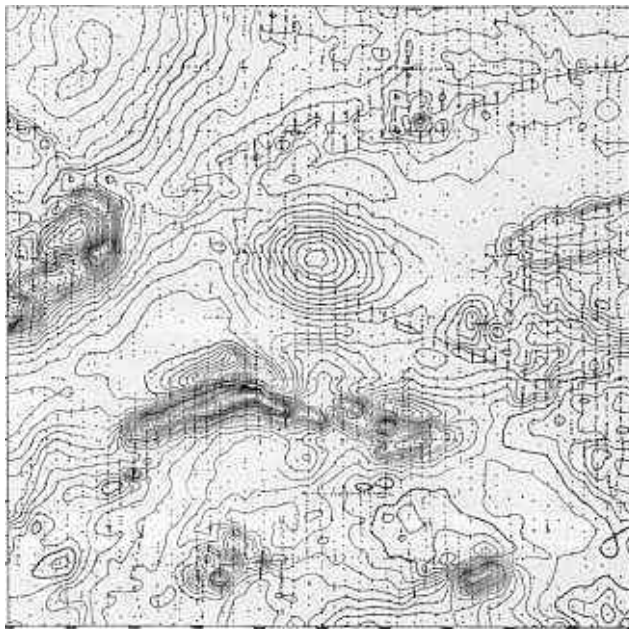
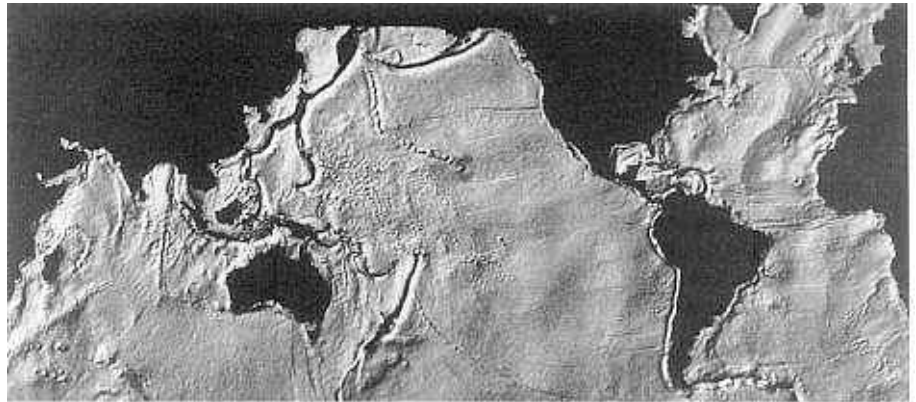


Fig. 8.11 (a) Contour map of magnetic total field anomaly over an area in northern England. (b) Image of the data shown in (a). The area is 100 km square.

8.2.5 Raster processing methods

Raster data of the kind discussed so far in this section contain a wealth of information from which geologists can make interpretations and combine them with other kinds of data for analysis in a GIS (Section 8.3). Although GIS methods provide awesome analytical power, for most applications visual interpretations are extremely effective. Figure 8.11 shows the improvement in visual quality that stems from converting contoured data to an image. In this case the data are aerial measurements of anomalies in the total magnetic field potential, converted for display from negative and positive values of nanotesla to an 8-bit (0–255) range of positive integers.

Although a great improvement over contoured data, the image in Fig. 8.11(b) suffers from two main problems. Human vision distinguishes between only 16 and 32 grey levels, so the 256 levels are not distinguishable. The image also lacks clues to 'depth' and therefore appears bland. These factors mask any high spatial-frequency features. Contrast stretching (Section 5.2) gives some improvement, but the full range of values for the magnetic anomaly only emerge graphically if the data are displayed in colour. This is most conveniently achieved by assigning the DN to a colour palette to produce a 'rainbow' range of colours, where low values are in blues and greens, intermediate values appear as greens and yellows, and high values show as orange to red and

magenta. The resulting pseudocolour image of the data in Fig. 8.1 is shown in Plate 8.1. Most magnetic and gravity data show deviations from the average potential field, and the anomalies may be positive or negative. Careful colour assignment can ensure that the zero anomaly coincides with the green to yellow transition in the rainbow spectrum, perceptually the most distinct boundary to us, so that 'cold' colours represent negative anomalies and 'warm' colours associate with positive.

Although pseudocolour renditions express more fully the data's information content, colour masks high-frequency spatial features in the data (Section 2.2). A lack of clues to depth also hides spatial detail. The most important clues to depth in our visualizing topography are shadows, as seen in all aerial and satellite imagery. Simulating solar shading of nonimage raster data uses a variety of algorithms, the simplest of which are the directional filters discussed in Section 5.3, which mimic the effect of a constant Sun elevation from the eight principal points of the compass. More sophisticated algorithms allow various illumination angles and types of lighting to be selected. Figure 8.12 shows the dramatic realism that results from such a transform of a DEM. You should see valleys and ridges realistically, but turn the page through 180 degrees and the illusion 'flips' to an inverse of topography (Section 2.5).

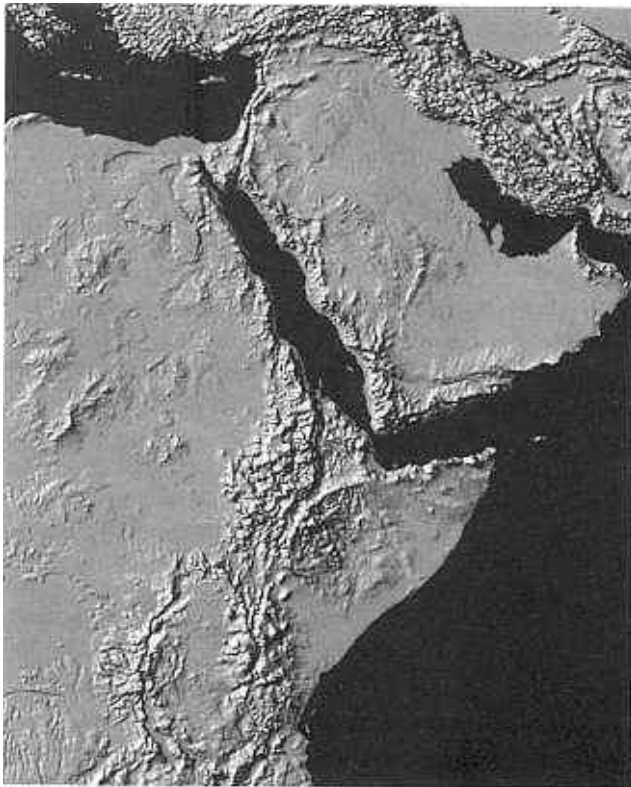


Fig. 8.12 Digital elevation data (GTOPO30) of East Africa filtered to simulate the shadowing by solar illumination from the north-west.

Raster renditions of any spatial variable are in effect maps of the 'topography' of the variable, and can be enhanced just as easily by simulated shading to tune an image to human vision. Figure 8.13 shows shaded images of magnetic 'relief' incorporating illumination from two directions at right angles. They show how high-frequency spatial features with different orientations can be highlighted separately. By combining oblique 'illumination' with pseudocolour, using multiplicative or ISH methods (Section 5.2.2), the advantages of both result in graphic images, ideally suited to interpretation (Plate 8.2).

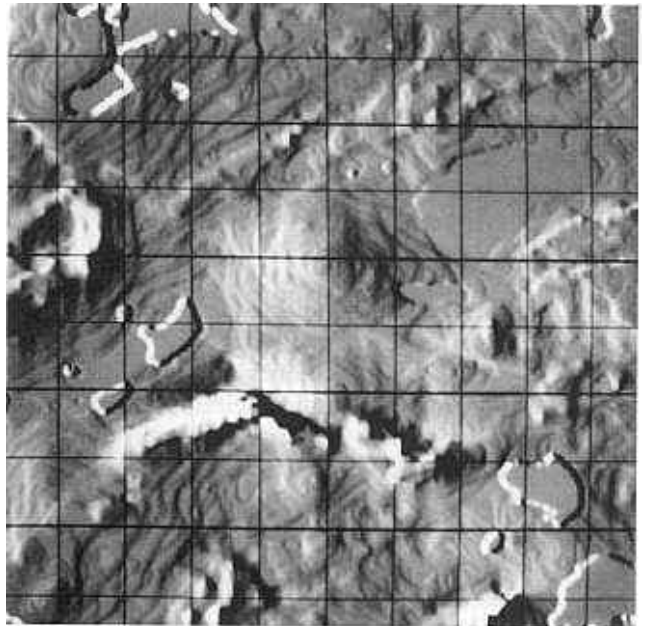
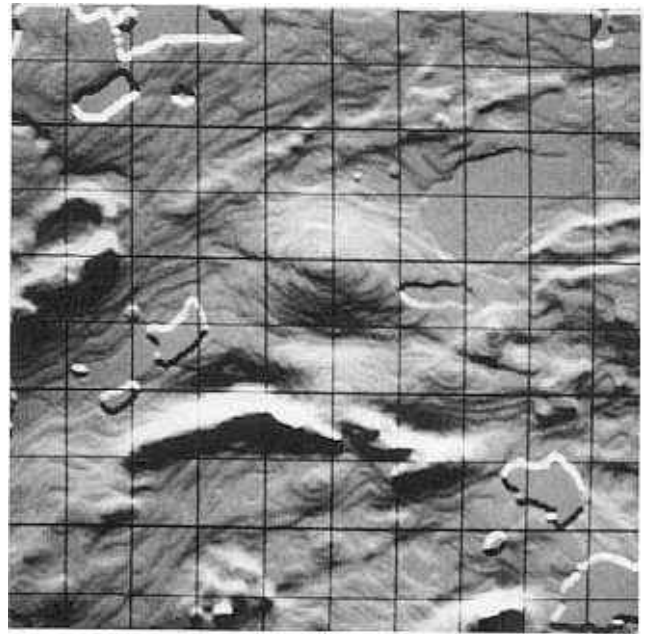


Fig. 8.13 Aeromagnetic data shown in Fig. 8.11(b) 'illuminated' from the north (a) and west (b).

As well as enhancement of gridded nonimage data, a variety of mathematical functions can extract information from them. These methods include differentiation to produce various derivatives and trend-surface analysis (Section 8.2.1). The magnitude of the first derivative of a spatial variable expresses the gradient of the topological surface. It is particularly useful in analysing DEMs by giving regional estimates of topographic slope. This can be useful in assessing slope stability and rates of runoff, as well as in geomorphological analysis. Some geological applications benefit from derivatives of geophysical data. Figure 8.14 shows an image of gravity data together with one expressing the slope of the gravitational field. The latter highlights sudden changes in gravitational potential, which may correspond with subsurface density boundaries.

The magnitude of the second derivative expresses the rate of change of gradient of a topological surface, and is thereby a measure of the variability or 'roughness' of a surface. Inflections in a surface by definition have a second derivative of zero, and so are defined uniquely. In force-field data the main element in contributing to high-frequency variations is the crystalline basement, in which magnetic susceptibility and density vary considerably in contrast to their relative uniformity in the overlying veneer of sediments. The more deeply buried basement features are the smoother their expression in gravity and magnetic data. In magnetic data, the relationship between anomaly patterns and the bodies responsible for them is often complex owing to bipolarity. Many magnetic anomalies resulting from a single body have both negative and positive features, and the source for the anomaly frequently resides beneath the line of its inflection. Consequently, second-derivative images of gravity and magnetic data express depth to basement qualitatively, and locate the sources of magnetic anomalies. Quantitative use of force-field data involves complex three-dimensional modelling, in which geologically reasonable bodies of different density and magnetic susceptibility are manipulated to simulate the observed gravitational or magnetic field anomaly patterns. They are beyond the scope of this book.

Another approach to analysing deeply buried features, especially from gravity data, is trend surface analysis (Section 8.2.1), which fits polynomial surfaces to the original point data. A first-order surface expresses the gross trends as a dipping plane, whereas higher-order surfaces assume more complex shapes to express regional patterns. The higher the order of trend-surface, the more closely it matches the actual variation in the data. However, even for 10th-order surfaces, the patterns are grossly smoothed compared with surfaces interpolated by spline or kriging functions, and represent the low-frequency spatial variations contained within the data. Low-frequency variations in gravity data stem from broad variations in

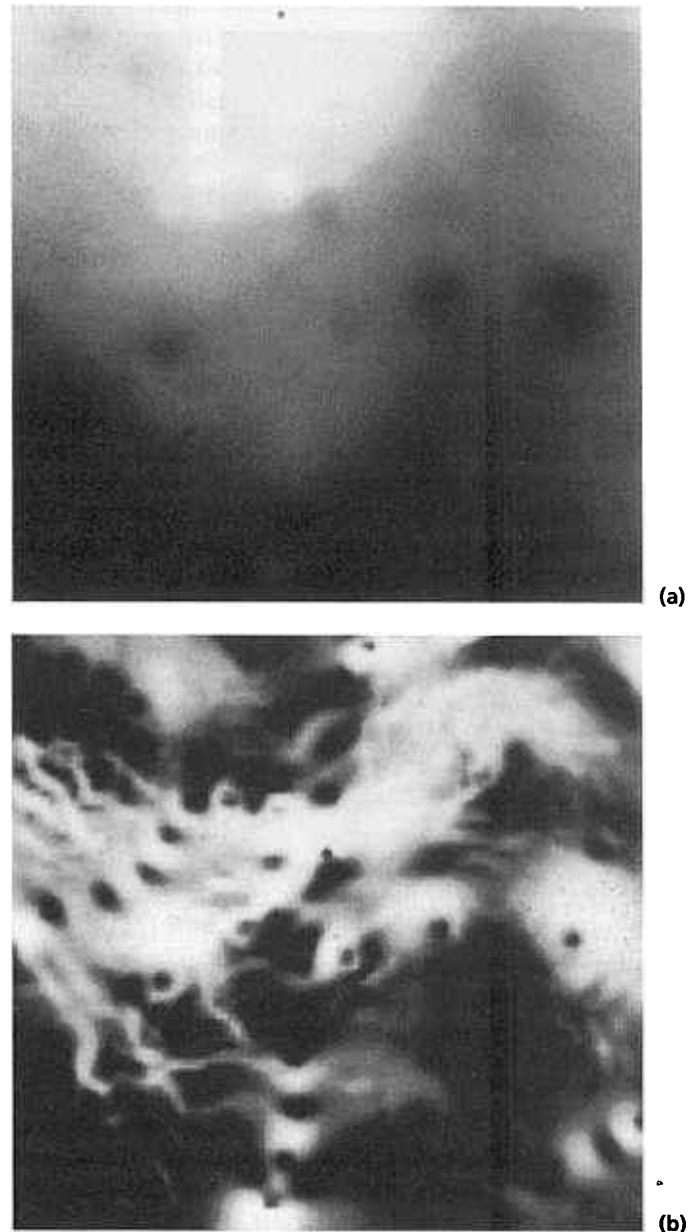


Fig. 8.14 (a) Image of gravitational potential in an area of Alaska, compared with (b) an image of the first derivative of the gravitational potential. Courtesy of Charles Trautwein, US Geological Survey, EROS Data Center.

the deep crust and upper mantle and from the relatively uniform but variably thick veneer of surface sedimentary cover. The higher frequencies stem from variations within near-surface basement, or perturbations in the depth to the cover-basement interface, often related to deep faulting. Subtracting the trend surface or regional field from the interpolated field topology results in a residual field, in which may reside important and previously indiscernible information about cover-basement relationships. Figure 8.15 shows the results of different orders of trend-surface and the associated residual fields

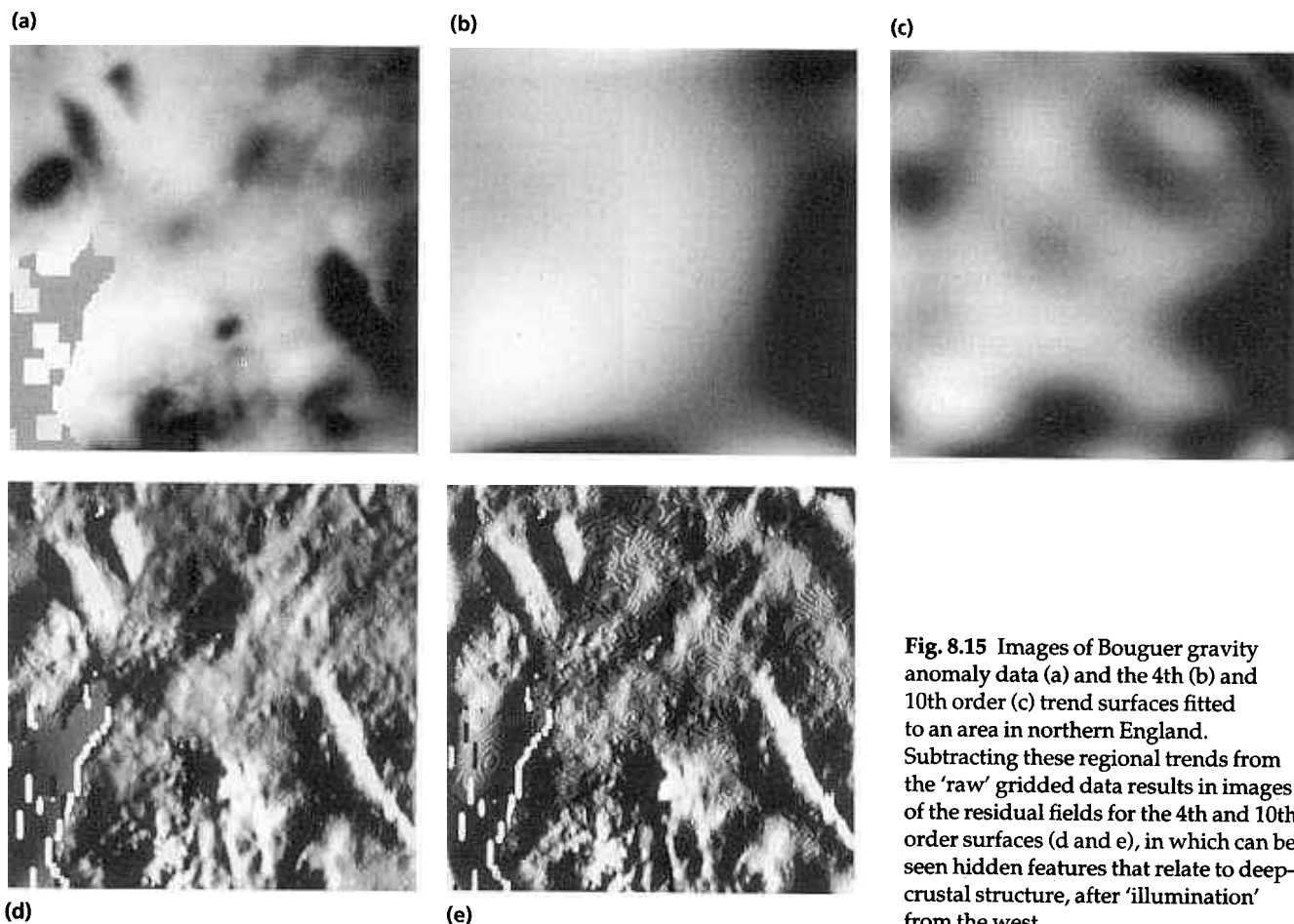


Fig. 8.15 Images of Bouguer gravity anomaly data (a) and the 4th (b) and 10th order (c) trend surfaces fitted to an area in northern England. Subtracting these regional trends from the 'raw' gridded data results in images of the residual fields for the 4th and 10th order surfaces (d and e), in which can be seen hidden features that relate to deep-crustal structure, after 'illumination' from the west.

from Bouguer gravity anomaly data for the area shown in Figs 8.11 and 8.13, and Plates 8.1 and 8.2(a).

8.2.6 Multivariate raster images

Enhanced images of single rasters are well suited to interpretation, using all the skills and intuition embodied in a trained geologist. Using them singly, however, when there are many different kinds of data, produces a welter of interpretations that need to be co-ordinated in some way. One approach is to transfer them to a GIS, where a variety of computer applications can assist in this task (Section 8.3), but this relies on mathematical means of seeking relationships. Human vision is by far the most flexible means of assembling connections between disparate sets of information, and it makes sense to exhaust the full range of possibilities presented by images before resorting to computer analysis.

Several methods combine multiple data sets as composite images, most relying on techniques covered in Chapter 5. The simplest of these uses three sets, i.e. the red, green and blue components of colour images. Seeking relationships between the sets relies on Young's theory of additive colour (Section 2.4). Where all three

variables are strongly correlated, the DN used in colour display are all similar, so that the result is a shade of grey, ranging from white to dark grey depending on whether all the values are high or low. If two of the variables are well correlated and the third is unrelated or anticorrelated, this is signified by the area being rendered in shades of yellow, magenta or cyan, depending on the colour assignment of the variables and their DN. Areas showing the three additive primary colours suggest that there is no significant correlation between any of the variables, and one is high whereas the other two have low values. Pastel and nonspectral colours result from partial correlation among the variables.

As RGB images assign the same weight to all three variables displayed, low information content in one variable tends to swamp that in the others, particularly if there are gross disparities in the spatial resolution of the data sets. Therefore, a simple rule is to combine data sets with approximately the same spatial resolution. Nor is it wise to combine data that have fundamentally different relationships to geological features. Consider an image made up from a Landsat band, gravity and magnetic data. By no stretch of the imagination can any meaningful correlation in an RGB image be expected, because the

properties being displayed relate to spectral reflectance, density and magnetic susceptibility, which have no relationship. Although all three may contain vital information on the distribution, say, of major faults, the resulting image inevitably will look like a pizza! A second rule therefore is to combine data in RGB form where there is some expectation of real correlation between the variables. For most geological applications, the unifying factor is likely to be different relationships to rock type.

A good example of variables to express in RGB order for lithological discrimination, if they are available, would be nickel, chromium and gravity anomaly data. As ultramafic rocks are rich in Ni and Cr and have high density, then areas underlain by peridotites would show as light grey tones on the image. Basalts generally have lower Ni/Cr ratios than ultramafic rocks and slightly lower densities. They might be revealed by greenish or cyan areas. Granites and sediments have low values for all three variables, and consequently would show as dark patches. Ironstones, with high density and low Cr and Ni, would be distinguished by blue signatures. Occurrences of nickel sulphide ores, being depleted in chromium but associated with dense ultramafic rocks, should be revealed as magenta or red areas.

Because much is now known about the affinities of different chemical elements for different lithologies, and their behaviour relative to one another in geological fractionation processes, geochemical data are ideally suited for simple RGB combination as a route to lithological discrimination. In particular the common association of restricted ranges of elements during different kinds of mineralization episode makes this approach extremely powerful in metal exploration. Some granites, with which a variety of pegmatite, skarn and hydrothermal deposits are frequently associated, consistently carry tin–tungsten–lithium–fluorine. In exploration for gold mineralization the ‘pathfinder’ elements arsenic–bismuth–antimony are sought routinely in drainage and soil surveys. However, high-resolution data from geochemical surveys are not often available outside of mining companies. Britain is particularly well endowed with public-domain data of this kind, and Plate 8.3 shows some highly informative results from them.

More widely available, particularly in arid and semiarid terrains, are the results of airborne gamma-radiation spectrometry (Section 3.7), which are often flown together with magnetic surveys. Gamma-radiation data relate to the abundances of radioactive isotopes of uranium, thorium and potassium in soils and rock. Fortunately, these three elements behave very differently in a great variety of geological processes, are found in a large range of concentrations and assume quite different relative abundances in a wide variety of common rock types. Although often noisy because of the blocking of radiation by vegetation and air, RGB images comprising U, Th

and K estimates from spectrometry are very colourful, the colours relating well to the surface distribution of different rocks and soils, as Plate 8.4 shows. Interpreting such images relies on both an understanding of additive colour theory and knowledge of the common relative abundances of the three elements in rocks. Granites usually contain all three elements in abundance, whereas basic and ultrabasic igneous rocks are strongly depleted. Sandstones are often depleted in both U and K, being dominated by quartz grains, but can have high Th contents locked in stable heavy minerals such as zircon and monazite. Clay-rich rocks generally contain high K contents bonded within clay minerals, and if deposited under reducing conditions generally have elevated U contents owing to precipitation and adsorption of the insoluble U^{6+} ion. They contain little Th owing to its retention in inert heavy minerals that rarely reach the low-energy environment associated with the deposition of clays. Limestones usually are depleted in all three, but if rich in organic remains appreciable U may have been precipitated from sea water or hydrothermal solutions because of the reducing conditions that accompany hydrocarbons.

Provided that the results can be understood easily, it is possible to combine information from up to six variables in an RGB image by using ratios. This is most suited to geochemical data, where ratioing is commonly used in expressing petrogenetic differences.

Despite the earlier warning about unwise combination of data in RGB form, meaningful fusion between variables with fundamentally unrelated structures but important geological content is feasible. The appropriate method is the intensity–saturation–hue (ISH) transform (Section 5.2.2). Although this was devised for the enhancement of RGB images by manipulation of the intensity and saturation derived from them, it is possible to substitute data of different kinds for the intensity and hue in the inverse (ISH to RGB) transform. The simplest application is using a single data set, such as aeromagnetic raster data. The raw data are used as the hue and a synthetic solar-shaded version incorporating high-frequency spatial features is used as the intensity, with saturation set to a constant (in the range 64–255) to ensure vivid colours in the result. This is another means of producing images similar to Plate 8.2.

Combining two data sets using the inverse ISH transform employs the strategy of using the data with the greatest spatial information content, often some kind of remotely sensed image, as the intensity input. This ensures that the spatial information is best rendered for visualization as brightness variations in the output colour image. Hue is controlled by the data with the greatest contribution to understanding lithological variation, such as gravity or geochemical data, so that the resulting colours are meaningful. Saturation is again set to a constant. Plate 8.5 gives an example of this approach.

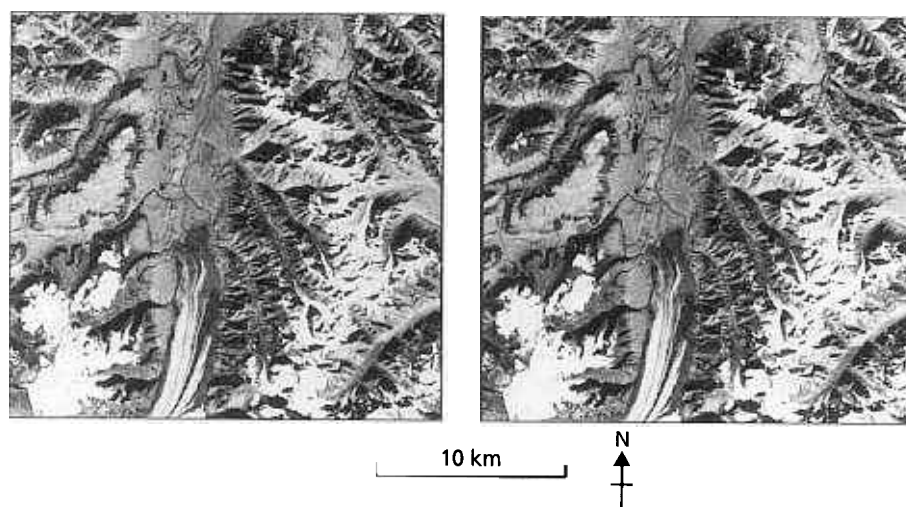


Fig. 8.16 The parallax in this stereopair of Landsat MSS images of Alaska was introduced using a digital elevation model. Courtesy of Charles Trautwein, US Geological Survey, EROS Data Center.

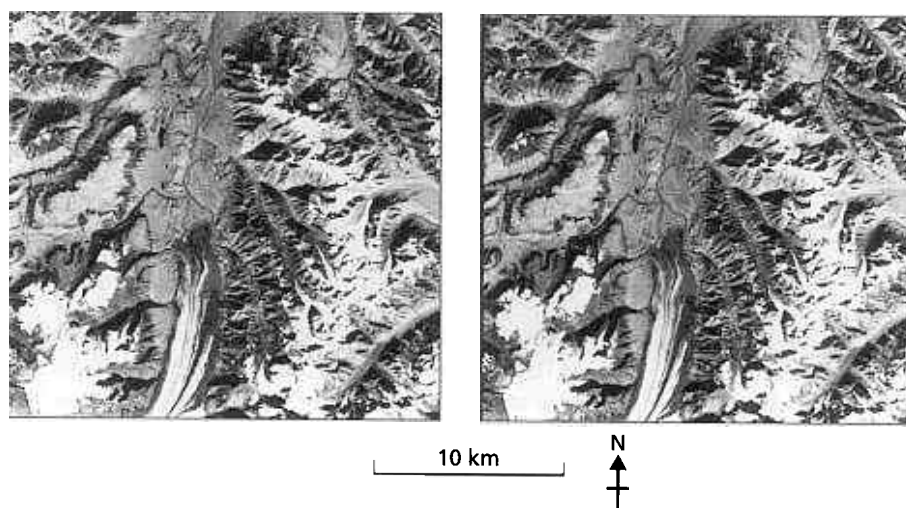


Fig. 8.17 The parallax in this stereopair of Landsat MSS images of Alaska was introduced using the gravity data shown in Fig. 8.14(a), and although the stereomodel looks odd, it helps resolve evidence from both the surface and subsurface. Courtesy of Charles Trautwein, US Geological Survey, EROS Data Center.

Because an RGB image comprising three variables can be converted to intensity, saturation and hue values using the ISH transform, the colour information, expressed by the saturation and hue, can be combined with another data set that is substituted for the intensity. Again, the best strategy is using data with potential for lithological discrimination as a source of colour and allowing intensity to be controlled by a variable containing high-frequency spatial information. Plate 8.6 incorporates side-illuminated aeromagnetic data as intensity and U, Th and K estimates from airborne radiometric data as saturation and hue for the same area as Plate 8.4. By careful use of ratios, this method allows the combination of up to seven variables, but at the expense of increasing the difficulty of interpretation. Usually such extreme methods use variables that are selected on the basis of a well understood model relating to the task in hand. For instance, epithermal gold mineralization is often associated with faults, generally results in clay-mineral alteration of host rocks, has elevated concentrations of

As-Bi-Sb and the hydrothermal fluids often deplete the hosts in a number of alkali metals. From that model a sophisticated use of the ISH method could be based on high-resolution imagery for intensity and a mixture of geochemical data and spectral ratios as the source of hue and saturation controls, but it may be less confusing to use GIS methods.

As parallax derived from stereopairs of images is a route to modelling variations in topographic elevation, so the inverse is possible. In the absence of stereoscopic images, which greatly improve the detection of geological structures and lithological boundaries, synthetic stereopairs can be generated from satellite images by adding parallax derived from a DEM. Figure 8.16 uses Landsat data and a DEM of part of Alaska. This approach is not limited to the use of elevation data, because the same procedure can convert any three-dimensional topology into parallax shifts in synthetic stereopairs. Figure 8.17 shows an example using gravity anomaly data to introduce relief into the Landsat image used in

Fig. 8.16. Neither is this method limited to a pair of data sets, but synthetic stereopairs can be produced for colour images that already may incorporate information from up to seven variables, as discussed above. Bizarre and potentially confusing as such a monumental fusion of data might seem, given careful blending of data to a particular geological model can produce images that fully exploit the geologist's powers of observation and mental synthesis (Chapter 9).

8.3 Data analysis in geographical information systems

Like different wavebands in multispectral remote-sensing data, the data layers residing in a GIS are registered, but with the additional advantage that the registration base is a geodetic co-ordinate system of some kind. A GIS is geocoded. Both raster and vector data, together with layers of attributes for each vector category, are likely to be present and used in a geologically orientated GIS. Unlike remotely sensed data, the contents of each file may not be related directly to the others. As will be shown, the purpose of a GIS is to derive new, topic- or task-orientated attributes and layers from interrogation of the original data. Consequently, large numbers of newly derived files are likely to be generated, many of which may be intermediate steps to some ultimate goal.

The most important general cautionary note is that using a GIS demands a far more ordered approach than does digital image processing. Its analytical potential should not simply be explored in the quest for visually stimulating images. It requires a carefully thought-out strategy that exploits information in the context of well-defined goals, and is constrained by the analytical functions that reside in the system itself. Without this discipline, the user undoubtedly will experience acute confusion and obtain results that have no significant meaning.

Geographic information systems use several different architectures. These involve two main types; those having a dominantly vector-based approach and those that are orientated to raster format. The first type mainly supports the handling of existing map data in digital form, and is more or less a kind of digital mapping system of interest to planners and cartographers. Raster-handling capabilities are essential when remotely sensed and other semicontinuous, spatial variables form an important proportion of the source data, as is undoubtedly the case in geological applications. Fortunately, there are now many GIS that combine both raster and vector capabilities. The link between the two kinds of function is the ease of conversion of vector files to raster format with various cell sizes, so that vector data can be treated as rasters. Obviously this means that several orders of mag-

nitude more storage and computing capacity is required than with a vector-based system, but the downward spiral of cost per unit of storage, processor speed and number-crunching power largely obviates these problems.

Much of the output from a GIS is in the form of maps that result from data fusion and analysis, and it is often necessary to display or print these to acceptable cartographic standards. For this, vector format has the distinct edge of sharp, accurate boundaries and ease of graphical modification. So, the most appropriate GIS for geology incorporate means of conversion from raster to vector format, at least for the simple boundaries and the categories within them that result from analysis.

The functions available in a GIS are numerous and often complex. Many were devised with the planner or manager in mind, not the geologist. The following sections summarize some of the most relevant functions, and more detailed accounts can be found in a number of specialist texts. Functions relevant to geological applications in a GIS break down to: retrieval and measurement (Section 8.3.1); overlay operations (Section 8.3.2); neighbourhood operations (Section 8.3.3); operations involving DEMs (Section 8.3.4); operations involving faults and fractures (Section 8.3.5).

8.3.1 Retrieval and measurement functions

Retrieval operations are the most simple functions in a GIS. They involve the search for specific attributes, values or ranges of values residing within the data available, and the transfer of the relevant points, lines, polygons or raster cells to a new file. Examples would be extracting all areas where granite outcrops, points where gold has been found, fold axes, areas of positive Bouguer gravity anomalies or where the abundance of a chemical element exceeds the threshold plus three times the average concentration (Section 8.2.3). More complex categories, such as all granites with high zirconium in stream sediments and low topographic relief also can be retrieved by specifying the data layers within which the relevant information lies. It is also possible to use retrieval operations to combine several classes as a means of simplifying information. An example would be combining all granites, adamellites, granodiorites, tonalites and monzonites of any age in an area as a general class of granitoids. The purpose of retrieval functions is to isolate specific features of interest, usually so that they may be combined with other information at a later stage. One possibility is setting all cells or polygons found to a value of 1.0 and other areas to zero. Multiplying other data sets by the retrieved file would lead to actual values only remaining for the retrieved class. This is a form of masking.

Measurement of the distance between points, the lengths of lines, perimeters and areas of polygons, the

areas occupied by raster cells conforming to some limits, and volumes of specified segments of a DEM are functions common to any GIS. Potential uses in geology may be as a first step in analysis of faults and fractures, assessing the relative importance of different rock types that may contribute toxic elements to drainage in a catchment area, measuring the areas and volumes of potential sources of aggregates, or even the road travel distances from one important locality to the next.

8.3.2 Overlay operations

Overlay operations can involve both arithmetic and logical approaches, and seek to combine information from two or more layers of data to derive another. Combinations involving arithmetic overlays can become extremely complex, so logical overlays are considered first. Both require considerable forethought and planning in the context of some model.

Logical operations involve the rules of simple Boolean algebra, which use the operators AND, OR, XOR and NOT to see whether a particular condition is true or false. These logical relationships are understood most easily using sets and Venn diagrams. Figure 8.18 shows a number of sets of attributes as Venn diagrams for a variety of logical statements, the shaded portions indicating the conditions where the statements are true. As an example of a simple but realistic model, consider a search for areas of potential hydrothermal mineralization in an area containing limestone. Limestones are prone to various types of hydrothermal mineralization, as they are attended by reducing conditions and are subject to easy solution by hydrothermal fluids. A common hydrothermal mineral in limestone country is fluorite (CaF_2). Given a geological map and stream-sediment fluorine determinations, an indicator of possible areas of mineralization would be derived by using the statement (limestone AND $\text{F} > 100$ ppm).

To illustrate the OR operator, consider a model for the location of useful wearing-course aggregate. In an area dominated by weak shales and sandstones there are small outcrops of basalts and slightly metamorphosed greywackes that would be ideal. The statement (basalt OR greywacke) would find all mapping units that are either basalt or greywacke, or both together.

For the NOT operator an illustration would be conditions for constructing a road. Areas of low slope would be ideal, but those occupied by peat deposits would be potentially disastrous. The statement (slope $< 2^\circ$ NOT peat) would show all gentle slopes except those occupied by peat.

The XOR operator is a little more complicated, and is equivalent to 'either A or B, but not both together'. This operator is rarely used in geological applications, but is useful in selecting areas for specific land-uses.

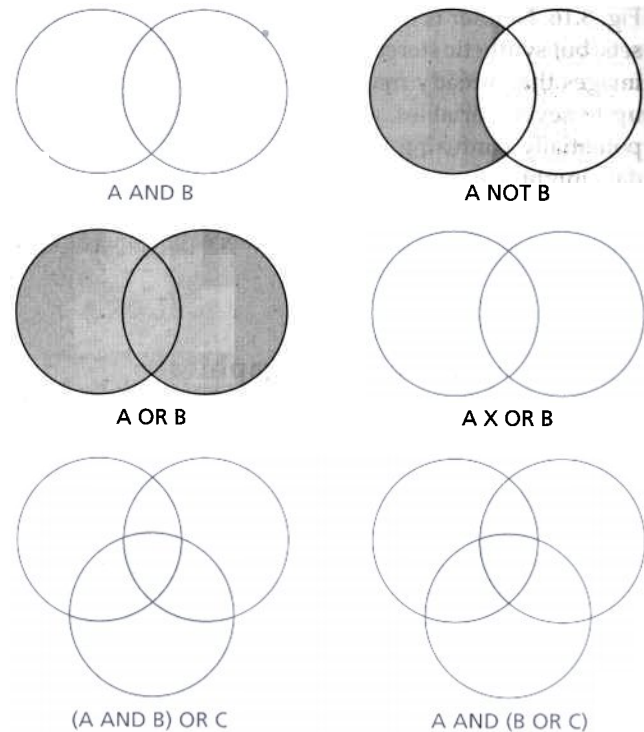


Fig. 8.18 Venn diagrams showing shaded areas that correspond to the Boolean logical statements AND, OR, XOR and NOT, and two more complex statements.

Logical operators can be applied to more than two attributes or conditions, but the operators must be in some form of priority. Figure 8.18 shows the different results of A and B OR C, the primary operator being parenthesized in the two cases. Consider the example of a search for mineralization in limestones. As well as fluorite, baryte (BaSO_4) is a common hydrothermal mineral in such terrains, and areas of limestone with either F or Ba anomalies might be prospective. The statement (limestone AND ($\text{F} > 100$ ppm OR $\text{Ba} > 200$ ppm)) would narrow the search, but a statement with the AND operator prioritized would also include areas with high F and Ba that are not associated with limestones, but perhaps with barren shales.

Results of logical overlay operations can be the end product of a search or the basis for masking, or they may be a step in a more complex sequence involving other functions.

Arithmetic overlay operations are appropriate when considering data that combine both attributes and variables, generally assembled to exploit some conceptual model. The general idea is to use the data and the model to rate or score areas or cells according to how well they satisfy the model. A good example is provided by targeting possible sites of porphyry-style copper mineralization. These always occur in association with intermediate to acid granitoid intrusions, so information

about their distribution is needed, perhaps combining areas where they crop out with measures of proximity to them (Section 8.3.3). As porphyry deposits occur near the top of such intrusions, seeking buried intrusions could prove important. Their low density and low magnetic susceptibility relative to country rocks produces negative anomalies on both gravity and magnetic maps. Copper concentration in drainage or soil geochemistry data is an obvious part of the model, but high copper can be associated with unmineralized basaltic rocks, so there needs to be some means of screening such false signals. Basalts contain far higher chromium concentrations than granitoids, so high Cu/Cr ratios should characterize mineralized granitoids. Porphyry mineralization is formed by the transport of metals in hot, watery fluids derived at a late-stage in the fractional crystallization of granitoid magma. Typically they result in alteration of feldspars to clay minerals, which have distinctive spectral signatures on remotely sensed images covering the SWIR. Porphyry deposits also contain large volumes of disseminated iron and copper sulphides, which are broken down to brightly coloured ferric hydroxides during weathering to give gossans. The model therefore should include data from remote sensing that highlight such alteration. Suitable data sets would be Landsat TM 7/5 and 3/1 band ratios.

Each criterion in the model must be assessed in terms of how the values of the variable involved can be divided into ranges that carry a score in favour of the presence of porphyry-copper mineralization. It is useful to set thresholds for some critical factors beyond which values definitely militate against a satisfactory conclusion, in order to exclude the relevant cells from the model. In developing the model, some criteria will be more encouraging than others, so the scoring often involves differently weighting the data sets used. The complex overlay operation then sums the weighted scores for each cell, resulting in a map that prioritizes the region for more detailed exploration and expenditure of funds. Virtually every kind of mineral deposit, hydrocarbon play or groundwater source could be expressed in such a form, so that the search is narrowed and rationalized.

8.3.3 Neighbourhood operations

Neighbourhood operations evaluate the characteristics of the area surrounding a specified location. One function is to count the number of specified features within a particular distance from a point, line or boundary. An example would be to count the number of mines within an economic trucking distance of every cell in an area as a means of deciding on possible sites for ore-processing plants that will serve several mines. That example would also use other considerations to highlight the most favourable sites.

A second function, that might help in this siting exercise would be measuring the distance of every cell from established transport routes. In the porphyry-copper example, this would be the means of establishing proximity to granite outcrops.

Many geological applications of GIS use proximity analysis in relation to various types of boundary. For example, many mineral deposits produced by movement of solutions through rock, such as hydrothermal, contact metasomatic and pegmatite types, are spatially related to faults, unconformities and contacts of igneous intrusions. There is a higher probability that they lie closer to such features than in homogeneous rock. Corridors or buffer zones of a specified width astride boundaries of these kinds open up more complicated, but appropriate overlay operations.

Many other neighbourhood operations rely on moving search areas that produce statistical summaries, including averages, totals, maximum and minimum values, and measures of diversity, for the immediate vicinity of points or cells in vector or raster data. Specialized texts will contain accounts of other neighbourhood operations that are more appropriate for land-use and other nongeological applications.

8.3.4 Operations involving DEMs

As well as estimations of topographic slope and simulation of solar shading using first derivative filters (Section 8.2.5), DEMs form the basis for extracting specific geomorphological features, such as lines of drainage, stream order, ridges and catchment boundaries. These operations involve a search outwards from a starting location using a specific decision rule. In seeking the path of water flow over a DEM the rule used might be to start at the highest elevation in the area and then move to the cell among the eight surrounding cells that has the lowest elevation. The operation would then be repeated until either it met the edge of the area, or a cell in a depression with no outlet. This would trace one drainage path. Starting again at the next highest elevation would trace another, and so on, until all topographically defined drainages had been defined. The connectivity between tributaries helps define the number of tributaries, and thus stream order, for all the drainages. Catchments above specified points can be defined by spreading only to adjacent cells with the same or higher elevation, until there are no adjacent higher cells, thereby defining watersheds. Such operations have great potential in hydrological and hydrogeological applications, where the land unit of interest is the catchment basin. Figure 8.19 shows watersheds and drainages paths that such analysis has extracted from a DEM of part of north-east Africa.

Although mainly of interest to planners, intervisibility functions define the areas in direct line of sight from

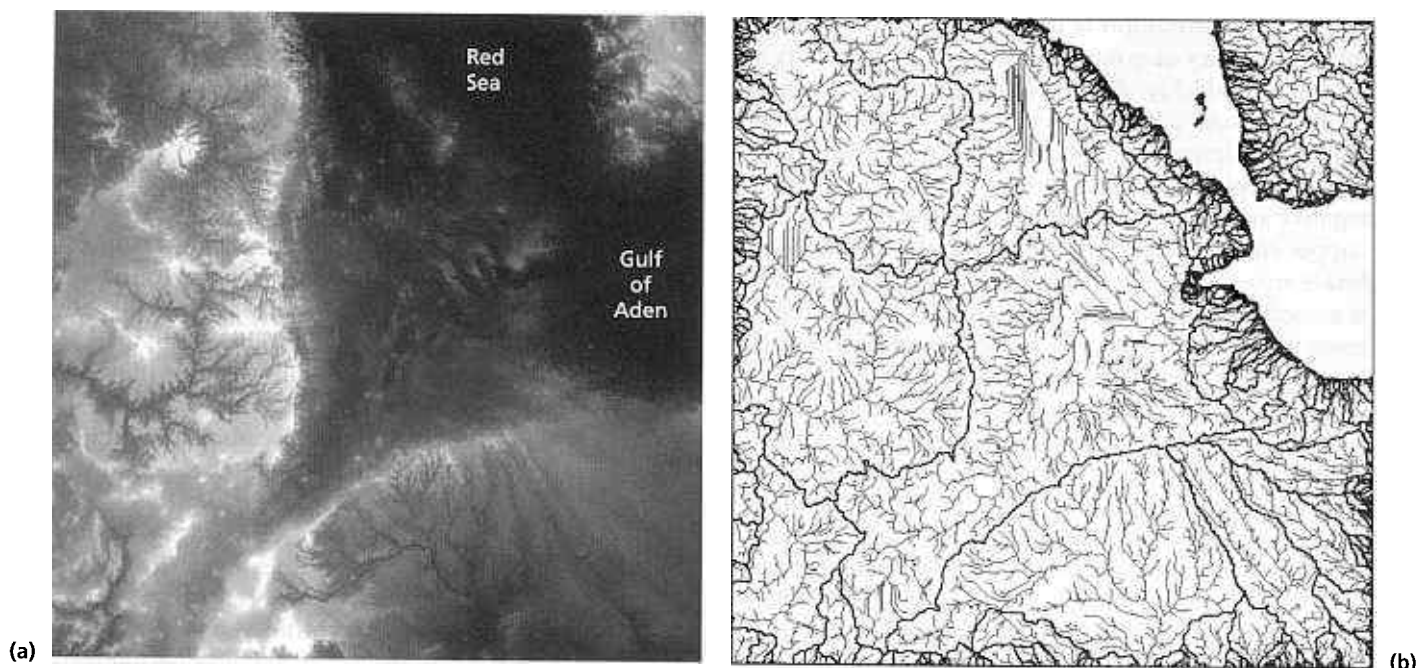


Fig. 8.19 Digital elevation model (GTOPO30) of central part of Fig. 8.12 (a), and (b) vectors for watersheds and drainage paths derived from it.

specified locations. This uses the DEM to find features that screen straight-line connections between the target location and parts of the surrounding landscape. This might be of use in satisfying the ever increasing environmental and amenity constraints placed on development of extractive industry and route orientation, with which geologists might easily become involved.

8.3.5 Operations involving faults and fractures

As well as the quest for hitherto undiscovered potato fields, lost cities and hidden oceans, it is tempting for the geologist to relax by drawing dense masses of straight lines on remotely sensed images. This often is supported by the misguided belief that all apparent linear connections between elements of terrain represent faults or fractures that surface mapping has missed. In truth, this once popular pastime owes much to imagination and the line-seeking predilection of human cortical cell assemblies (Section 2.1), for individuals rarely agree on all linear features observed on images. Nonetheless, suitably enhanced images do reveal many known and unsuspected real faults and fractures, the trick being to devise means of ascertaining their reality.

Now, the presence of fractures in the crust, particularly those involving dilatation, has considerable significance for the channelling of various kinds of fluids. Exactly what the fluid dynamics are in fractured rocks as opposed to those in porous and permeable media is not well known. What is abundantly clear is the strong

economic connotation of fracture-controlled mineral deposits, hydrocarbon migration paths and ground-water resources, as well as leakage potential from fractured waste-disposal sites and seismic risk in heavily faulted terrain. Hence the 'near obsession' of some remote sensing geologists with fracture analysis, and the growth of a not inconsiderable cottage industry. Geological remote sensors are sometimes regarded with a jaundiced eye in consequence of a welter of learned papers in the 1970s and 1980s, which rested on masses of tedious, web-like maps and largely meaningless rose diagrams indicating the length, density and azimuth of lines on images.

Given that it is possible to weed the illusory from the real, the digitization of mapped and interpreted faults and fractures in vector format now allows useful statistical information to be extracted. Software is available to define fracture spacing and density, the density of fracture intersections, spatial trends and a variety of more sophisticated measures. These derivatives can then be combined with other data layers using operations covered in Sections 8.3.1–8.3.3.

8.4 Concluding note

The use of non-image data in raster form and application of GIS methods to multivariate information is at the high end of image interpretation in terms of costs in time and money. Set against this are a number of mitigating

factors. Much non-image data initially is generated for limited uses, and if they fail to deliver the 'goods' they are often discarded, despite the frequently high costs of acquisition. For many of them, ingenuity can easily devise other applications. Indeed, hidden in their original form is important information with unexpected connotations. Recasting them in digital form, either as rasters or in vector format, helps reveal these surprises, particularly by using routine enhancement techniques. Multiple uses divide survey and data-processing costs among several projects. The basic element in GIS methods is integration of data, and there is no reason why this should not lead to integration of seemingly unrelated projects. As an example, it requires no stretch of the imagination to see that drainage geochemistry can serve mineral exploration, monitoring risks to humans, livestock and environments, and in some cases may help explain features of natural vegetation. All three have other common links, such as soil type.

By establishing an ethos of using spatial data bases and GIS methods for all aspects of the environment in its broadest sense, public administrations from the local, through national to regional and global levels can foster closer links between all the disciplines involved. Each has contributions to make to all the others. Social, economic and environmental problems are not unrelated in reality. Treating different problems separately often leads to unsuspected difficulties in another field. So such a fusion of effort as well as data can chart a course that takes as its starting point the interconnectedness of all things.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file Menu.htm in the Imint folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

On the Menu page, select the links Additional Images: Geophysical, Geochemical Data or Digital Elevation, where you will find examples of images based on a variety of nonimage data.

Further reading

- Anfiloff, V. & Luyendyk, A. (1986) Production of pixel maps of airborne magnetic data for Australia, with examples for the Roper River 1 : 1 000 000 sheet. *Exploration Geophysics* 17, 113–117.
- Aronoff, S. (1989) *Geographic Information Systems: a Management Perspective*. WDL Publications, Ottawa.
- Arvidson, R.E., Guinness, E.A., Strebeck, J.W., Davies, G.F. & Schultz, K.J. (1982) Image processing applied to gravity and topographic data covering the continental United States. *EOS (Transactions of the American Geophysical Union)* 63, 261–265.

- Batson, R.M., Edwards, K. & Eliason, E.M. (1976) stereo and Landsat pictures. *Photogrammetric Engineering and Remote Sensing* 42, 1279–1284.
- Bolivar, S.L., Balog, S.H. & Weaver, T.A. (1982) Resonance characterization for uranium mineralization in the Montrose 1° × 2° Quadrangle, Colorado. *Proceedings of the 5th Annual Uranium Seminar, New York*, pp. 37–48. American Institute of Mining Metallurgy and Petroleum Engineering, New York.
- Bonham-Carter, G.F., Agterberg, F.P. & Wright, D.F. (1988) Integration of geological data sets for gold exploration in Nova Scotia. *Photogrammetric Engineering and Remote Sensing* 54, 1585–1592.
- Burrough, P.A. (1986) *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford University Press, Oxford.
- Campbell, A.N., Hollister, V.F., Dutta, R.V. & Hart, P.E. (1982) Recognition of a hidden mineral deposit by an artificial intelligence program. *Science* 217, 927–928.
- Drury, S.A. & Walker, A.S.D. (1987) Display and enhancement of gridded aeromagnetic data of the Solway Basin. *International Journal of Remote Sensing* 8, 1433–1444.
- Drury, S.A. & Walker, A.S.D. (1991) The use of geophysical images in subsurface investigations of the Solway Basin, England. *Surveys in Geophysics* 12, 565–581.
- Edwards, K. & Davis, P.A. (1994) The use of intensity–hue–saturation transformation for producing colour-shaded relief maps. *Photogrammetric Engineering and Remote Sensing* 60, 1369–1374.
- Gibson, P. (1992) Evaluation of digitally processed geophysical data sets for the analysis of geological features in Northern Ireland. *International Journal of Remote Sensing* 14, 161–170.
- Graham, D.F. & Bonham-Carter, G.F. (1993) Airborne radiometric data: a tool for reconnaissance geological mapping using a GIS. *Photogrammetric Engineering and Remote Sensing* 59, 1243–1249.
- Guinness, E.A.R.E., Arvidson, C.E., Leff, M.H., Edwards, K. & Bindshadler, D.L. (1983) Digital image processing applied to geophysical and geochemical data for Southern Missouri. *Economic Geology* 78, 654–663.
- Harrington, H.J., Simpson, C.J. & Moore, R.F. (1982) Analysis of continental structures using a digital terrain model (DTM) of Australia. *BMR Journal of Australian Geology and Geophysics* 7, 68–72.
- Hastings, D.A. (1983) Synthesis of geophysical data with space-acquired imagery: a review. *Advances in Space Research* 3, 157–168.
- Hoffman, P.F. (1987) Continental transform tectonics: Great Slave Lake shear zone (~ 1.9 Ga), north-west Canada. *Geology* 15, 785–788.
- Jenson, S.K. & Domingue, J.O. (1988) Extracting topographic structure from digital elevation data for geographical information system analysis. *Photogrammetric Engineering and Remote Sensing* 54, 1593–1600.
- Junkin, B.G. (1982) Development of three-dimensional spatial displays using a geographically based information system. *Photogrammetric Engineering and Remote Sensing* 48, 577–586.
- Karpuz, M.R., Roberts, D., Olesen, O., Gabrielsen, R.H. & Herrevold, T. (1993) Applications of multiple data sets to structural studies on Varanger Peninsula, Northern Norway. *International Journal of Remote Sensing* 14, 979–1003.
- Kowalik, W.S. & Glenn, W.E. (1987) Image processing of aeromagnetic data and integration with Landsat images for improved structural interpretation. *Geophysics* 52, 875–884.

factors. Much non-image data initially is generated for limited uses, and if they fail to deliver the 'goods' they are often discarded, despite the frequently high costs of acquisition. For many of them, ingenuity can easily devise other applications. Indeed, hidden in their original form is important information with unexpected connotations. Recasting them in digital form, either as rasters or in vector format, helps reveal these surprises, particularly by using routine enhancement techniques. Multiple uses divide survey and data-processing costs among several projects. The basic element in GIS methods is integration of data, and there is no reason why this should not lead to integration of seemingly unrelated projects. As an example, it requires no stretch of the imagination to see that drainage geochemistry can serve mineral exploration, monitoring risks to humans, livestock and environments, and in some cases may help explain features of natural vegetation. All three have other common links, such as soil type.

By establishing an ethos of using spatial data bases and GIS methods for all aspects of the environment in its broadest sense, public administrations from the local, through national to regional and global levels can foster closer links between all the disciplines involved. Each has contributions to make to all the others. Social, economic and environmental problems are not unrelated in reality. Treating different problems separately often leads to unsuspected difficulties in another field. So such a fusion of effort as well as data can chart a course that takes as its starting point the interconnectedness of all things.

Associated resources on the CD-ROM

You can access the resources by opening the HTML file Menu.htm in the Imint folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

On the Menu page, select the links Additional Images: Geophysical, Geochemical Data or Digital Elevation, where you will find examples of images based on a variety of nonimage data.

Further reading

Anfiloff, V. & Luyendyk, A. (1986) Production of pixel maps of airborne magnetic data for Australia, with examples for the Roper River 1 : 1 000 000 sheet. *Exploration Geophysics* 17, 113–117.

Aronoff, S. (1989) *Geographic Information Systems: a Management Perspective*. WDL Publications, Ottawa.

Arvidson, R.E., Guinness, E.A., Strebeck, J.W., Davies, G.F. & Schultz, K.J. (1982) Image processing applied to gravity and topographic data covering the continental United States. *EOS (Transactions of the American Geophysical Union)* 63, 261–265.

Batson, R.M., Edwards, K. & Eliason, E.M. (1976) Synthetic stereo and Landsat pictures. *Photogrammetric Engineering and Remote Sensing* 42, 1279–1284.

Bolivar, S.L., Balog, S.H. & Weaver, T.A. (1982) Resource characterization for uranium mineralization in the Montrose 1° × 2° Quadrangle, Colorado. *Proceedings of the 5th Annual Uranium Seminar, New York*, pp. 37–48. American Institute of Mining Metallurgy and Petroleum Engineering, New York.

Bonham-Carter, G.F., Agterberg, F.P. & Wright, D.F. (1988) Integration of geological data sets for gold exploration in Nova Scotia. *Photogrammetric Engineering and Remote Sensing* 54, 1585–1592.

Burrough, P.A. (1986) *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford University Press, Oxford.

Campbell, A.N., Hollister, V.F., Dutta, R.V. & Hart, P.E. (1982) Recognition of a hidden mineral deposit by an artificial intelligence program. *Science* 217, 927–928.

Drury, S.A. & Walker, A.S.D. (1987) Display and enhancement of gridded aeromagnetic data of the Solway Basin. *International Journal of Remote Sensing* 8, 1433–1444.

Drury, S.A. & Walker, A.S.D. (1991) The use of geophysical images in subsurface investigations of the Solway Basin, England. *Surveys in Geophysics* 12, 565–581.

Edwards, K. & Davis, P.A. (1994) The use of intensity–hue–saturation transformation for producing colour-shaded relief maps. *Photogrammetric Engineering and Remote Sensing* 60, 1369–1374.

Gibson, P. (1992) Evaluation of digitally processed geophysical data sets for the analysis of geological features in Northern Ireland. *International Journal of Remote Sensing* 14, 161–170.

Graham, D.F. & Bonham-Carter, G.F. (1993) Airborne radiometric data: a tool for reconnaissance geological mapping using a GIS. *Photogrammetric Engineering and Remote Sensing* 59, 1243–1249.

Guinness, E.A.R.E., Arvidson, C.E., Leff, M.H., Edwards, K. & Bindschadler, D.L. (1983) Digital image processing applied to geophysical and geochemical data for Southern Missouri. *Economic Geology* 78, 654–663.

Harrington, H.J., Simpson, C.J. & Moore, R.F. (1982) Analysis of continental structures using a digital terrain model (DTM) of Australia. *BMR Journal of Australian Geology and Geophysics* 7, 68–72.

Hastings, D.A. (1983) Synthesis of geophysical data with space-acquired imagery: a review. *Advances in Space Research* 3, 157–168.

Hoffman, P.F. (1987) Continental transform tectonics: Great Slave Lake shear zone (~ 1.9 Ga), north-west Canada. *Geology* 15, 785–788.

Jenson, S.K. & Domingue, J.O. (1988) Extracting topographic structure from digital elevation data for geographical information system analysis. *Photogrammetric Engineering and Remote Sensing* 54, 1593–1600.

Junkin, B.G. (1982) Development of three-dimensional spatial displays using a geographically based information system. *Photogrammetric Engineering and Remote Sensing* 48, 577–586.

Karpuz, M.R., Roberts, D., Olesen, O., Gabrielsen, R.H. & Herrevold, T. (1993) Applications of multiple data sets to structural studies on Varanger Peninsula, Northern Norway. *International Journal of Remote Sensing* 14, 979–1003.

Kowalik, W.S. & Glenn, W.E. (1987) Image processing of aeromagnetic data and integration with Landsat images for improved structural interpretation. *Geophysics* 52, 875–884.

- Lee, M.K., Pharaoh, T.C. & Soper, N.J. (1990) Structural trends in central Britain from images of gravity and aeromagnetic fields. *Journal of the Geological Society of London* **147**, 241–258.
- Majumdar, T.J., Mohanty, K.K. & Srivastava, A.K. (1998) On the utilization of ERS-1 altimeter data for offshore oil exploration. *International Journal of Remote Sensing* **19**, 1953–1968.
- Marble, D.F. & Peuquet, D.J. (1983) Geographic information systems and remote sensing. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 923–958. American Society of Photogrammetry, Falls Church, Virginia.
- McGuffie, B.A., Johnson, L.F., Alley, R.E. & Lang, H.R. (1989) IGIS: computer-aided photogeologic mapping with image-processing, graphics and CAD/CAM capabilities. *Geobyte* **October**, 8–14.
- McMahon, M.J. & North, C.P. (1993) Three-dimensional integration of remotely sensed geological data: a methodology for petroleum exploration. *Photogrammetric Engineering and Remote Sensing* **59**, 1251–1256.
- O'Sullivan, K.N. (1983) The role of image processing in mineral exploration: why stop at Landsat? *Advances in Space Research* **3**, 169–171.
- O'Sullivan, K.N. (1986) Computer enhancement of Landsat. Magnetic and other regional data publications. *Volume 2, Geology and Exploration, 13th CMMI Congress, Singapore*, pp. 1–10.
- Paterson, N.R. & Reeves, C.V. (1985) Application of gravity and magnetic surveys: the state of the art in 1985. *Geophysics* **50**, 2558–2594.
- Reeves, C.R. (1985) Airborne geophysics for geological mapping and regional exploration. *ITC Journal* **3**, 147–161.
- Salvi, S. (1995) Analysis and interpretation of Landsat synthetic stereopairs for the detection of active fault zones in the Abruzzi region (central Italy). *Remote Sensing of Environment* **53**, 153–163.
- Seemuller, W.W. (1989) The extraction of ordered vector drainage networks from elevation data. *Computer Vision, Graphics Image Processing* **47**, 45–58.
- Wadge, G., Young, P.A.V. & Mason, D.C. (1992) Simulation of geological processes using an expert system. *Journal of the Geological Society of London* **149**, 455–463.

9 Geological Applications of Image Data

Most applications of remote sensing in geology involve the delineation of structures and the discrimination of different rock and soil types, often in spite of variations in natural vegetation cover but sometimes exploiting them. This chapter examines the contexts and potential roles for remote sensing in the day-to-day activities of professional geologists. Much can be learned from individual case studies, but such is the range of possibilities that to illustrate them well would require a book on each application area in its own right. Any project has its own internal logic, geared to the nature of the terrain, the availability of image data, costs and efficiency, and the specific objectives of the investigation. Many case studies skate over the generalities or omit them for the sake of brevity, assuming that readers have some knowledge of geological image interpretation, and even for reasons of commercial confidentiality.

Selecting a remote sensing strategy involves considering how geological models comprise a number of general components, and seeking means whereby information from images or other graphic data can assist in assembling these components. Chapter 9 covers several topics in which geologists play a crucial role, examines the kinds of information that they need to contribute in helping to achieve the objectives of a particular programme, and suggests the kinds of data and the way that they can be handled appropriately to provide this information. It is not a 'cookbook' but a guide to orientating the methods of remote sensing to the observational and analytical skills that have long been the geologist's stock in trade.

Can remote sensing supplant field observations? The answer is a resounding 'No!', for a number of reasons. First, many of the features that are essential in building up a geological picture, such as petrographic texture, minor tectonic and sedimentary structures, and detailed stratigraphic and petrogenetic relationships are far too small to be resolved by any conceivable remote sensing instrument. Second, truly fresh rock is rarely, if ever exposed, being covered by weathering patinas, soil and vegetation of different types. The hammer, hand lens and ultimately the microscope and geochemical analytical techniques ferret out the mineralogical and textural features that geologists use in classifying rocks and understanding their genesis. Third, even if fresh rocks were exposed completely, detailed knowledge of the spectra of rock-forming and accessory minerals reveals that few can be identified and separated uniquely, nor can their proportions be assessed precisely because of variations in the strength of their associated spectral features in pure form. In the reflected region detectable minerals

are restricted mainly to those carrying ferric, hydroxyl and carbonate components, together with differences in overall reflectivity that stem from a multitude of factors. The thermally emitted region is more promising because of the shifts in spectral features associated with variations in silicon and aluminium co-ordination with oxygen and other molecular features, which give information on rock-forming minerals. Although giving some guide to rock composition, overlaps and different intensities of spectral features still mean that multispectral thermal data are unlikely ever to provide a comprehensive mapping capability. Radar data provide information on the weathering and erosional textures of rocks, which help determine variations in surface roughness, and the presence of corner reflectors smaller than the method's resolution dimensions. Although helpful in some cases by providing another dimension to image interpretation, neither is radar an all-purpose geological mapper in its own right.

So information provided by remote sensing has to be verified and amplified by field work, in much the same way as the subsurface information provided by geophysical surveys needs to be interpreted in the light of surface observations, and ultimately by the drill. However, remote sensing views the surface in a detailed map-like form, providing a synopsis of areas that are far larger than can be encompassed by a ground view and much faster than field mapping. It reveals features that are difficult to map at the surface and allows field observations to be extrapolated into unvisited areas. In this regard remote sensing reduces the amount of field work needed to achieve project objectives to, and also directs those investigations to areas containing the most critical evidence. All this helps increase a geologist's efficiency. By venturing into nonvisible parts of the spectrum and through digital image processing, remote sensing increases the chances of making discoveries that would be overlooked in conventional field work. It amplifies the geologist's powers of observation and detection.

Remote sensing provides a kit of tools that a geologist can deploy as and where necessary. Making a selection from this toolbox involves several considerations:

1 Application definition. What geological features and processes relate to the application? Is it possible to extract information about them from image data?

2 Resolution and scale. What is the minimum spatial resolution needed to detect the required geological information? Over what range of scales do the geological features and processes manifest themselves? In general, the finer the resolution the larger is the scale of

presentation. The maximum scale needed is often a guide to the minimum resolution size. If the requirement involves a maximum scale of 1 : 1 million then there is little point in using data with a resolution of 10 m or less. As Equation (2.1) reveals, what the eye can see in an image depends on the scale and viewing distance, and there are definite limits. A project involving products at 1 : 10 000 scale needs a resolution less than 5 m.

3 Spectral coverage. Is multispectral, single band or panchromatic imagery required, and what regions of the spectrum are appropriate for the features being sought? The answers depend on whether the main emphasis is lithological discrimination, mapping geological structures or evaluating terrain. The first hinges on the degree to which rocks and the soils derived from them are free of vegetation cover, and needs selection of wavebands that mineralogically determined spectral features affect. Structural mapping and terrain analysis depend to a large extent on the interpretation of high-frequency topographic features. So, our better perception of fine detail in grey-tone images (Section 2.2) than in colour images often means that single-band or panchromatic images are more appropriate than false-colour composites, particularly if they can be viewed stereoptically. Structural geology and lithological mapping are complementary, however, and some tectonic features can be inferred only by their displacement, or relationship to lithological boundaries, when colour images can be useful. Similarly, geomorphology involves considering the relationship between physical processes of landscape development and the effects of vegetation cover, when false-colour images are the best means of providing information on plant distribution. For some applications nonimage data are appropriate, such as topographic elevation, potential-field or geochemical data in a form suitable for use in image processing or GIS operations.

4 Data processing. Chapters 5 and 8 discussed a wide range of possible manipulations used in enhancing or extracting the information contained in images. Because all involve time and therefore cost, and in some cases produce new images that require varying depths of insight into what they reveal, a choice of techniques is important in planning a project. Some are aimed at improving the visual qualities of images, such as contrast stretching and edge enhancement, others at expressing hidden information and relationships between different data sets, such as ratioing and principal component analysis. A further suite extracts information from many spatially related data sets, as in classification and GIS operations.

More mundane considerations are whether the data required are indeed available, how much they would cost, and if they have acceptable cloud cover and atmospheric haze. An important point concerns the time of year for data acquisition, as emphasized in Section 5.6.3

and Plate 5.10. This involves the presence of living vegetation (sometimes this is a distinct aid, sometimes a hindrance, depending on the local climatic conditions), varying soil-moisture content and the Sun angle at the time of overpass. At moderate to high latitudes the winter Sun is lower in the sky than in the summer, so that the shadows of muted topographic features are more apparent. In a few applications that involve short-lived phenomena, such as studies of volcanoes, changing river courses and environmental changes associated with resource-based industry, repetitive cover is essential. For most geological applications, however, a single image with favourable atmospheric and vegetation conditions is often all that is needed. Moreover, the date of the image is frequently immaterial (this can be financially important, as data from early years are often cheaper than new data). There are, of course, areas that are so plagued by perpetual cloud cover that reflected and thermal images rarely reveal the surface. Whatever the application in such an area, only radar remote sensing provides any assistance to the geologist.

9.1 Geomorphology

The study of landforms and landscape involves description, classification and analysis of the evolution of the planet's surface. Although sometimes seen as a marginalized topic within geology, geomorphology is of fundamental importance because it is one of the driving forces of biological evolution, and controls habitability. In landscape studies the element of detail is essential, for landforms are enormously complex on all scales. The information base has to match reality, and images at different scales are the only realistic base for full interpretation.

Remotely sensed images lend themselves ideally to geomorphological studies at any required scale and area of coverage. As well as specific topographic and hydrological features, images can provide details on vegetation type and cover, on which landscape evolution depends together with climatic variations, underlying geology and time. The images in Chapter 4 provide a rich compendium of geomorphological examples, as well as illustrating most aspects of geology.

Until the widespread accessibility of satellite images the science had drawn in on itself, after the megascopic ideas of Davis fell into disfavour. It focused on the small scale, the short-term and the study of process, becoming almost divorced from geology. This trend coincided with the rise of ideas on systems analysis and the application of statistical methods, to which detailed information on small areas lent itself. The landscape became idealized as elements linked by flows of mass and energy. Such an approach found itself incapable of extrapolation to

large features and time spans of geological magnitude. It was an exclusively 'bottom upwards' approach. Studies of other planetary bodies, such as the Moon, Mars and satellites of the Outer Planets based themselves entirely on images of successively increasing scale, forcing a 'top downwards' approach and much different methods, which largely relied on stratigraphic rules of observation. Elements concerned with planetary evolution began to emerge. As well as discovering features of gigantic dimensions, this approach needs considerable ingenuity to interpret the processes involved. Planetology found an echo in geomorphology, with the focus shifting once again to grand theories.

Climatic geomorphology builds on Davis' stageist theory of evolution. It is based on the notion that today's relief-forming mechanisms vary according to zoned climatic conditions, and this relationship can be extrapolated to the past. Much of the terrestrial landscape is seen as being inherited from processes during past climatic periods. As climate currently is in transition from a major warming only 10 000 years ago this seems inescapable. Before the end of the last glacial period the world was a much different place from today for about 100 000 years, and imposed features that have not yet come to balance with modern climatic conditions. Indeed, the last 2–4 million years have been a period of great climatic instability, and the pace of succeeding glacial and interglacial periods has a far higher frequency than climatic shifts in earlier, more 'normal' geological times. Elements of a steady-state, pre-Pliocene set of landscapes still play an underlying role in setting today's geomorphological scene. But climate is not the only dynamic in the evolution of the Earth's surface.

Structural geomorphology focuses on the fundamental control of geology and tectonics over landscape, partly by different resistances of rocks to erosion and partly because of active vertical movements of the crust through different mechanisms. Climate plays a role but its effects are conditioned by rocks and their movements. A unity of the two strands of terrestrial planetology arose from the growing recognition that rapid erosion by glaciers and powerful rivers unloads the lithosphere, thereby allowing uplift. Removing mass, mainly from valleys, unloads the lithosphere beneath the remaining peaks and higher ground so that they continually rise, thereby generating mountain ranges of Himalayan and Alpine magnitude. This recognition stems from the discovery that much of the rapid Himalayan uplift took place in the Pliocene and Pleistocene, whereas the underlying tectonic forces that generated thick, buoyant continental crust were initiated tens of million years before. The most rapid uplift was delayed until rapid erosion became possible when climate changed so that glaciers and the meltwater derived from them could sculpt the landscape at an accelerated rate.

Clearly, geomorphology is scale-dependent and unification is only possible by studies at all scales from that showing minor gullies to continent-wide features. From a lower limit of around 1 : 10 000 supported by aerial photography and the highest resolution satellite data, remotely sensed data presents a continuum of scales down to 1 : 10 million or less through platforms such as Landsat and SPOT, and the coarse-resolution meteorological satellites. Even terrain features on the submetre scale, as expressed by surface roughness, are amenable to radar imaging.

Although most landscapes change at rates that are too slow to be monitored within the 60-year history of systematic remote sensing, some processes are quick enough for change detection to be possible. These include aeolian, coastal, glacial, volcanic, lacustrine, mass-wasting and some fluvial processes. Judicious selection of time series of an area, if available, can help monitor the dynamics involved in landscape change.

Because landscapes are three-dimensional, stereoscopy is of great importance in geomorphology. Where organized agriculture camouflages topographic features with patchwork patterns (Fig. 4.48) it is essential. Gradually the SPOT programme, and later stereoscopic systems, are providing global stereoscopic coverage that previously was possible only through large-scale aerial photographs, which do not allow continuity of three-dimensional visualization. Although stereoscopic Large Format Camera images covered several million square kilometres, sadly the deployment of this low-cost instrument on the Space Shuttle was limited to just one mission in 1984.

In rigorous statistical analyses of landform stereoscopic images include extraneous clutter as a result of variations in the reflective properties of the surface at a wide range of spatial frequencies. Moreover, proceeding from a mental stereomodel to actual measurements of relief is a tedious process. Accurate expressions of the variation in topographic elevation through digital elevation models form the base for quantitative landform analysis (Section 8.2.4). Digital elevation models may be derived from topographic contour maps, by automated parallax estimation from digital stereoscopic imagery and more recently by interferometric radar. Such DEMs enable the scale-dependence of terrain features to be monitored, amounts and rates of mass transfer to be estimated and drainage channels and watersheds to be extracted automatically. The entire continental surface is covered at about 1 km resolution by the GTOPO30 DEM, to be supplanted at about 100 m by the SRTM interferometric-radar system (Chapters 7 and 8) planned for completion by 2001–2 by JPL-NIMA.

Geomorphological mapping poses considerable problems as a result of the many different approaches that are in vogue and the sheer complexity of landforms. All

mapping involves simplifying the complex continuum of reality through symbols, categorization and areal division—a sort of data reduction in its own right. Three possible general approaches to division of landforms are: purely as functions of shape and pattern; according to interpretation of the genetic mechanisms involved; and by recognizing landforms of different ages using simple stratigraphic rules. It is difficult, if not impossible, to avoid welding the three approaches together, because evidence for all three manifests itself to the interpreter on images. Drainage systems provide ready examples of the confusion. These can be categorized purely on form (Fig. 4.3), but the forms always have some underlying controls, often due to geological factors, and may have been induced by an earlier fluvial episode subsequently incised into a newly rising block of crust or as a result of land subsidence or rising sea-level. Changes in erosional base level, both falling and rising, dominate those parts of the Northern Hemisphere once covered by continental ice-caps, and all continents have responded in terms of erosion and deposition to the global sea-level rise that followed retreat of those ice-caps. Depending on tectonic activity, such eustatic changes are accompanied by isostatic changes in base-level. It is quite common in any climatic zone to find meandering drainages that developed on an early planation surface, which are now incised into rugged, rising mountains (Fig. 4.15).

Only the most grossly generalized geomorphic categories are widely agreed upon. In practice no two geomorphological maps produced by different interpreters are comparable in detail, and few correspond except in a very general way. Producing maps is a graveyard for many a geomorphologist with a bureaucratic outlook! Essentially, terrain maps cater for specific needs, not for all. Some have attempted to map *geomorphological units*, which are genetically homogeneous landforms produced by specific constructional or destructional processes. All land-forming processes overlap, however, and many landforms owe their features to several processes of different age. A more productive line of attack relies on a hierarchy of forms depending on scale. In such a *multitiered approach*, one tier of large-scale map units disappears or blurs as scale and perceptual detail decrease, to be replaced by a succession of tiers at increasingly smaller scales. In this way it is possible to make links between manifestations of process and evolution from the scale of the smallest element of interest to those that characterize entire continents. This engenders a unification between understanding the detailed dynamics of currently active systems, through the interplay between fundamentally different types of erosional and constructional processes in an evolutionary context, to the grandiose relationships between surface and internal processes that involve the lithosphere as a whole, areas the size of continents and time spans of tens, hundreds and perhaps even thousands

of million years. Trying to achieve such a symbolic relationship, rich in the dynamics involved is the challenge posed to geomorphologists by planetologists working on extraterrestrial bodies.

Even if this systematization were to be achieved, there is still the likelihood that geomorphological maps will be made by experts for experts. There is no problem or harm in the context of academic research, but it often seems that remotely sensed images themselves are the most useful geomorphological maps, because all the information is there for producing thematic maps of any type for any purpose. However, practical decisions have to be made and boundaries drawn. Consequently geomorphologists are often called on to make interpretations and maps to assist a specific social, environmental or economic programme. Co-ordinating image interpretation involves isolating the elements of the landscape that are relevant to the programme's objectives. Geomorphology with an end other than human curiosity is often termed *terrain analysis*. It contributes to environmental and resource management, land-use and route planning, military strategy, and engineering applications (Section 9.4), increasingly through GIS combined with other types of information, such as variations in vegetation cover, soil type, geology and hydrology.

Chapters 4 and 7 contain many useful guidelines for geomorphological image interpretation, as do some of the articles in the further reading for this chapter and modern textbooks on geomorphology. Images from the reflected region are quite sufficient for most geomorphological purposes, and it is rarely impossible to achieve substantial results with single-band images alone. However, false colour images, if affordable, can add significantly to interpretations. As the dominant features on images are geomorphological in origin, little enhancement is necessary, except for cosmetic improvement by contrast stretching and edge enhancement.

Thermal data and radar images increase the breadth of detectable features, with radar adding information on microrelief that forms a direct link to field observations. Where cloud cover dominates, as in equatorial regions, radar sometimes forms the only source of information about terrain, because of its all-weather capabilities. Attempts have been made in such obscured regions to produce radar mosaics, such as Project RADAM in Brazil. Of the existing orbital SAR instruments, that carried by ERS-1 is inappropriate for geomorphology in all but areas of low relief because of its steep depression angle and the dominance of images in rugged terrain by layover (Fig. 3.34). The JERS-1 SAR has a less steep depression angle and produces images as useful as those from SIR-A, as can Radarsat. Radar has another important attribute to offer geomorphologists, in the form of shallow penetration of very dry sand and ice. Figure 7.4 gives a dramatic example of the potential for discovering

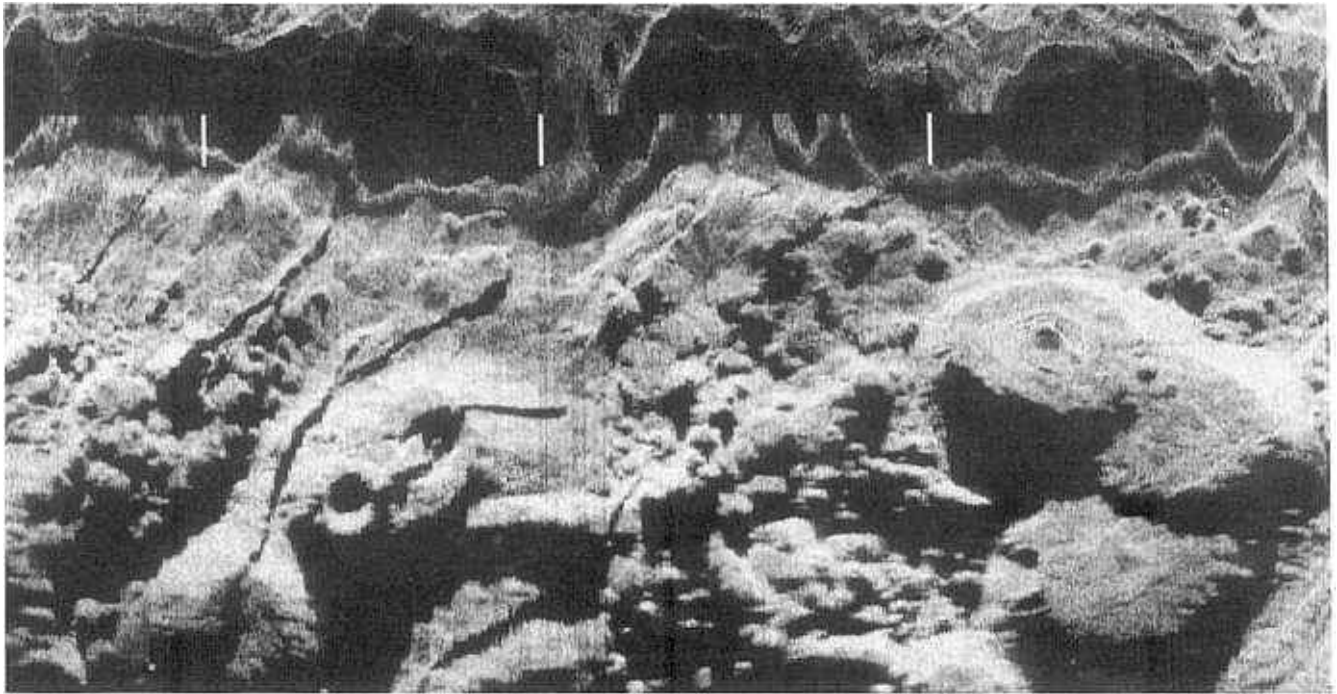


Fig. 9.1 Sidescan sonograph of part of the axial zone of the Mid-Atlantic Ridge. Sonar images are similar to those produced by radar, looking sideways from the platform track and responding to timed acoustic returns. They are both shadowed and vary in tone according to surface roughness. In this case, the dark upper strip is directly beneath the platform—a submersible. Prominent on the image are

flat-topped shield volcanoes at right and a conical volcano with a funnel-shaped crater at left. Separating these features is a hummocky terrain of huge pillow-like flows along the axial rift. A fault scarp is prominent at left. The image is 3 by 4 km. Courtesy of Bramley Murton, Institute of Oceanographic Sciences, UK, and Roger Searle, University of Durham, UK.

completely hidden geomorphological information using SAR in hyperarid areas. The potential global coverage of JERS-1, ERS-1 and RADARSAT promises an upsurge in geomorphological discovery.

It is often forgotten that the 70% of Earth's surface covered by oceans has landforms too. Conventional bathymetry is limited in its contribution to mapping the morphology of the ocean floor through its restriction to shipping lanes. The Seasat and ERS-1 radar altimetry data, through measurement of sea-surface elevation contain information relevant to water depth through gravitational effects (Fig. 8.9). Side-scanning sonar reveals many unsuspected features of the sea bed (Fig. 9.1) that is beginning to revolutionize understanding.

9.2 Geological mapping

Although *Image Interpretation in Geology* is about extracting geological information from images, the main responsibility of geologists is rationalizing this information in the form of graphic maps. How such maps are designed and produced is mainly outside the book's scope, but is covered using practical examples in the CD-ROM exercises. Essentially, mapping consists of

defining boundaries between rock units of different type, expressing their disposition relative to each other and to time, and delineating the tectonic structures about which they are disposed. Geological maps are two-dimensional expressions of four dimensions—three spatial dimensions and time, and incorporate clues to Earth history and the interplay between internal and surface processes.

Remote sensing cannot provide this information unless it is blended with field work and laboratory investigations. As emphasized many times, image data do not directly provide conventional divisions between rock types, which are and will depend on mineralogy, texture in hand specimen and thin section, and geochemistry. There are two main, general approaches to mapping from images; one based on *image units* (see Figs 6.11 and 6.12) the other on *lithofacies units*.

Image units combine the textural and pattern attributes of the surface with the way it interacts with EM radiation in terms of albedo or spectral characteristics expressed on false-colour images. It is a more or less objective approach to mapping, and implies no genetic connotations. For extraterrestrial surfaces, with little if any chance of ground verification, such an approach is inevitable. The relative ages of the units derive from simple stratigraphic rules, such as discordant relationships. In

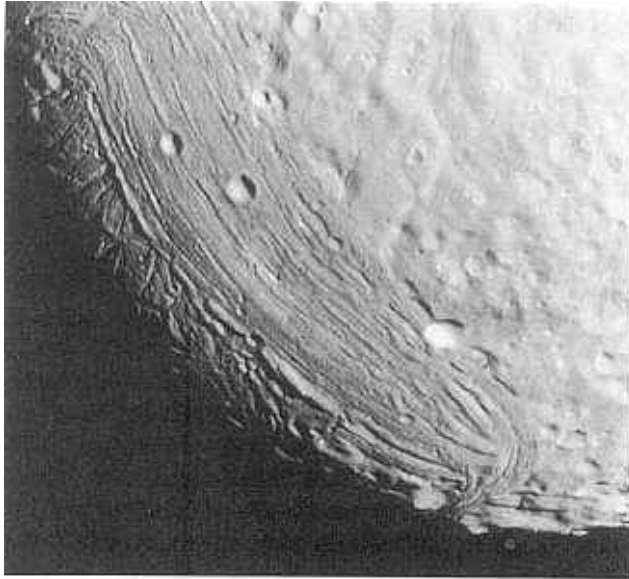
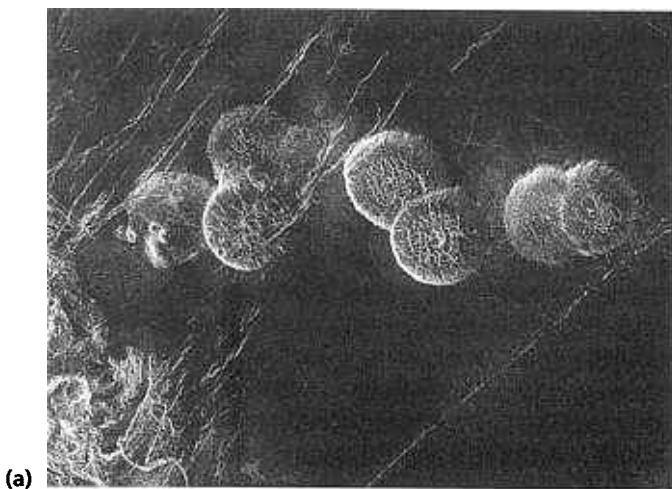


Fig. 9.2 Voyager-2 image of Miranda, a moon of Uranus, showing a grooved terrain next to an older cratered terrain that has been partially resurfaced to mute the craters. Younger impact craters clearly show as sharp circular features. Courtesy of the Jet Propulsion Laboratory, Pasadena, USA.

studies of many planetary bodies variations in the distribution of impact craters become useful (Fig. 9.2). This is based on the observation, initially from the Moon, that the older the surface the greater the chance of craters having formed, the rate of impacting having decreased with time throughout the Solar System. Other such signs of relative age include the masking of features by transported debris, such as ejecta from impacts or aeolian dust.

Occasionally there is sufficient detail to recognize general shapes that are reminiscent of processes on the Earth produced by the flow of fluids of different viscosity, including water, ice, lavas and landslips (Fig. 9.3). Using image units has the appeal of being definable in terms of the available data, but different units emerge from using different data (Fig. 6.11). More importantly, such units are of little use to other geologists, and the main requirement of a geological mapping programme is to lay out information that has a practical use, without the user having to go through the rigmarole of understanding the particular means of subdivision.

In terrestrial mapping, there are few cases where field data or earlier geological maps are not available. Consequently there are often opportunities to assign lithologies to recognizable image units, extrapolate from known to unvisited areas and to trace structural features over broad areas. Even in cases of completely unknown terrain—increasingly rare—image interpretation can proceed directly to a reconnaissance map of conventional type. Enough will be known from adjoining, mapped terrains to make an educated guess as to the likely rocks and structures there, backed up by spectral knowledge and the appearance of different rocks in climatically similar terrains. Geologists are unlikely to find many surprises in terms of rock composition that require a complete revision of terrestrial petrogenesis, such as vast flows of sulphur lava or methane ice that occur on other planetary bodies. Geologists make mistakes during image interpretation of new areas without field data, but where there is no information virtually any new observation is valuable, and new ideas help guide field work to test them out. Armed with experience of the ways in which



(a)

Fig. 9.3 (a) Magellan SAR image of the Alpha Region region of Venus showing volcanic domes, around 25 kilometres across, formed by extrusion of viscous magma. The curious fracture patterns on the domes suggest extrusion after the formation of a solid skin. Courtesy of Jet Propulsion Laboratory,



Pasadena, USA. (b) Viking image of Valles Marineris on Mars, in which the branching form of some of its side valleys suggest the former action of flowing water on this now dry planet. Courtesy of Jet Propulsion Laboratory, Pasadena, USA.

geological boundaries are modified by topographic relief it should be possible to predict where hidden boundaries trend and reappear.

In general, the vast majority of geological boundaries reflect either roughly parallel compositional layering or discordances cutting across that layering, such as faults, igneous contacts or unconformities (Figs 4.20, 4.24a and 7.10). Recognition of layer-parallel features on images, conforming to the relationship between dip and strike of planar fabrics and relief is of primary importance, for this enables folding and angular discordances to be discovered, evaluated and mapped. Occasionally, layers with distinctive spectral properties or texture (Plates 6.5 and 7.2) serve as marker horizons that help in more accurate delineation of fold structures and relative displacements on faults (Plate 9.1). Although the amount of information that permits such photogeological mapping varies with vegetation and soil cover, this more subjective approach based on inferred genetic relationships leads to results that have a predictive content of immediate use to any geologist. It forms the bones on which more geological flesh can be sculpted as investigations proceed. Achieving this often depends on experimentation with digital image processing to reveal the best combination of data, and access to a variety of renditions helps glean all the clues available.

Of all possibilities, stereoscopic viewing is undoubtedly the most important aid to grasping the relative disposition of rock types and the dip of surfaces in relation to relief. It shows clearly the attitude of layers and their thickness, and using simple instruments they can be quantified accurately. The law of Vs (Section 4.3.1) on single images is a good guide to the direction of dip. Topographic elevation data used together with this help quantify the dip by using the simple three-point solution for elevations on an assumed plane. Elevation values at many points on outcropping boundaries as they cross a DEM provide structure contours, after appropriate surface fitting. Alternatively, a DEM can introduce parallax into a single image, so that when viewed stereoptically with the original, three-dimensional relief is combined with the image features (Fig. 8.16). Because of the way in which low-angle planar features follow topography closely, they are extremely difficult to detect and map on images taken from directly above, even if they are stereoscopic. Powerful software allows the use of DEMs in creating oblique views that incorporate perspective, from any azimuth and altitude above the surface. 'Draping' images on such views simulates the 'Swiss hammer' approach to low-angle faults and unconformities (Plate 9.2).

Many existing geological maps were constructed from field work and aerial-photograph interpretation, often with generalization about lithological variation, or to suit a specific purpose that limited the number of divisions. Very often the 'nose-to-the-ground' approach prevented

the recognition of large or subtle features. Important divisions were frequently missed, simply because they were invisible. Mapping huge tracts at small scales (1 : 100 000 and smaller) relies on traverses where the terrain permits easy and rapid access, rather than by systematic visits to every outcrop and the following of boundaries. Inevitably, such reconnaissance maps omit much detail, but a general framework does exist. Remotely sensed data helps revise these maps at larger scale. Digital cartographic methods open up new possibilities. Scanning an existing colour geological map into red, green and blue separates and then registering it to an image, allows a geologist to combine the two by using techniques such as the HIS transform (Section 5.2.2) (Plate 9.3). Digitizing the boundaries from existing maps, either automatically or manually, is an alternative. Merging the boundaries with images provides guides to the distribution of lithologies and structures. Either approach allows these vectors to be edited and new ones to be added as the first step in conversion from paper to digital maps. Comprehensive interpretation of all conceivable and accessible geological attributes in an area is difficult to express graphically in a single paper map without distracting clutter and confusion. The high costs of mastering and printing lithographic paper maps generally mean that much of this information, although available, is rarely accessible except by consulting archives. Storing all interpreted attributes as vector lines, polygons and points in a digital mapping system lays the basis for entirely new styles of map production. Trying to jam as much information as possible on to a conventionally published map results in compromises, omissions and simplifications. Instead the objective becomes to generate maps that extract detail pertinent to specific applications from the mapping data base as 'one-offs'. Moreover, all information is available for further integrated analysis using GIS approaches.

Geological mapping aims at understanding the subsurface as well as surface geological features, so it benefits from using geophysical data, such as those from magnetic and gravity surveys (Section 8.2.2, Figs 8.1–8.15). Not all means of subdividing rocks are provided by field or image observations. Plate 8.3 demonstrates that gamma-radiometric surveys contain a wealth of information related to lithological variation. Plate 9.4 shows the potential of merging such data with more spatially 'clean' remotely sensed images.

9.3 Exploration

Although the guidelines for finding mineral deposits, hydrocarbon reservoirs or water are geological, exploration itself is fundamentally an economic activity. No anomaly is sought merely out of curiosity. Finding abnormally high concentrations of metals, oil or water

beneath the surface is not the end of the story either, for they do not necessarily constitute resources. To become economic entities, they must satisfy a wide variety of criteria. They must have a ready market, and be extractable and refineable to a useful product at a total cost that is less than their intrinsic value, thereby ensuring profit. As well as the multifarious technicalities associated with extracting the sought-for products from beneath the ground and refining them, there are the purely logistic aspects of access and supply of services. Unlike any other commodities physical-resource production inevitably depletes the assets of the operating organization; most physical resources are nonrenewable. Industries based on renewable resources and manufacturing either increase assets or at least maintain them at a stable level. Intrinsic values or market prices of physical resources fluctuate in a largely unpredictable fashion, part of which is linked to changes in the available supplies. Consequently, successful exploration can rebound on the market, as was the case with the discovery of large petroleum reserves outside the ambit of the OPEC cartel, which since 1973 had attempted to maintain oil prices at a high level. Together with other economic forces and measures to conserve energy in the 1980s, these discoveries led to a depression in the world price of oil that continues to this day, and means that exploration is continually subject to stringent limits. In the early 1990s these limits were so rigid that major oil companies began to disband the large teams of remote sensors that had been assembled in their exploration departments since the mid-1970s. Much the same trend affected virtually every company involved with metals and lower value resources, partly owing to gluts on the market stemming from new discoveries and more cost-effective extraction technologies, and partly to increasingly unpredictable markets. The concept of what constitutes a resource or a reserve changes with time. Water is in a unique position, for it is a vital requirement for survival and for the development of the largely agricultural economies of the so-called 'Third World', as well as for the smooth running of the social and industrial systems of the developed countries.

More than any other industrial activity, exploitation of and exploration for physical resources are bound up with risks to invested capital. The attraction of capital depends on these higher risks being offset by the opportunity for a greater return on investment. No-one will seek even gold if the required capital can be more profitably and safely invested elsewhere. Exploration is akin to research and development—it carries no guarantee of profit in its own right, and certainly its successful outcome after evaluation is not immediate. Most resource operations require a lead time of up to 15 years before offsetting the costs of exploration against income from production. However, the depletion of assets demands that exploration continues, except in periods

of fundamental economic depression, such as that following the Wall Street Crash of 1928.

This is not the place for a full account of the non-geological factors relating directly and indirectly to exploration. Suffice it to say that it centres on decisions about what to look for, where to concentrate the search, a minimum quantity of the resource in question for it to generate profit by exploitation, a minimum unit value, depending on the concentration of the commodity being sought and its unit price, and a minimum rate of profit. This in a nutshell, is an *exploration objective*. Note that only the location of the search area involves geological considerations, and they are mixed up with other factors, such as varying fiscal conditions, accessibility and communications, and political stability. To a variable extent exploration involves a gamble—at the outset there is only a set of optimistic notions, particularly when most of the deposits of the 'trip-over' variety have long been discovered. The greatest risk therefore is that of 'gambler's ruin', or the loss of capital through a string of unsuccessful ventures. Avoiding this disagreeable situation means estimating the probability of satisfying the exploration objective, and assigning funds in proportion to the risk of failure throughout the whole operation. Exploration therefore is conducted in stages, which gives every opportunity to proceed or stop according to the accumulation of information, not as a blind gambler would. Each stage involves a review of information gained and the degree to which the risk of not satisfying the exploration objective has changed. The expected value of the information from the next stage is set against the cost of getting it. When the value exceeds the cost exploration continues. When cost exceeds the value of expected information then it stops. This cumulative process amounts to setting lost opportunities against the chances of lost investment. Figure 9.4 shows the kinds of method available to metal-orientated exploration, and the stages at which they are often deployed, in relation to growing costs. Table 9.1 shows some approximate figures for costs and efficiencies of various exploration techniques involved in metal exploration. Although outdated in terms of the actual costs, it shows the relative expense and efficiency of each method.

The low cost and high efficiency of remote sensing makes it a favoured method early in exploration. It both reduces early costs and helps narrow the focus of later stages. This is important because narrowing the area of search at the end of each exploration stage goes hand in hand with reducing the degree of risk to investment. As well as helping geological reconnaissance, remote sensing is useful in *targeting* smaller areas for more detailed and costly follow-up. Moreover, its use does not end with the preliminary stage. The interpretation can be updated in the light of other data, and imagery forms an ideal base upon which to co-ordinate this later information through a GIS approach.

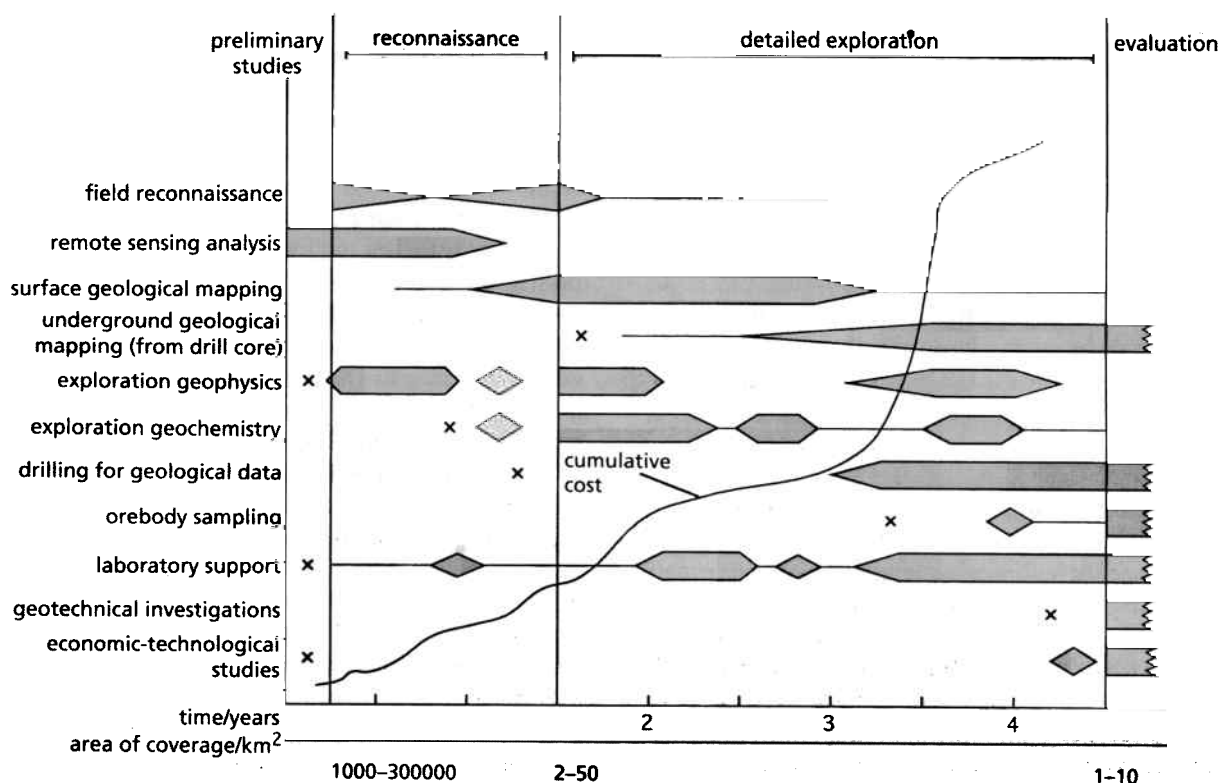


Fig. 9.4 Chart showing the general stages involved in a mineral exploration programme. The graph of costs is only an indication of the way cumulative costs mount.

Table 9.1 Approximate costs and average efficiencies of some exploration methods (1988 data)

| Methods | Cost (US dollars) | Efficiency (km ² day ⁻¹) |
|---------------------------------------|-----------------------|---|
| <i>Preliminaries</i> | | |
| Satellite remote sensing | 0.02 km ⁻² | >10 ⁶ |
| Interpretation and map | 0.7 km ⁻² | 10 ⁴ |
| Airborne remote sensing | 10 km ⁻² | 500 |
| Interpretation and map | 5 km ⁻² | 50 |
| <i>Airborne geophysics</i> | | |
| Magnetic and electromagnetic | 25 km ⁻¹ | 500 |
| Interpretation and map | 10 km ⁻² | 25 |
| Literature search | 250 day ⁻¹ | - |
| <i>Field studies</i> | | |
| Geological reconnaissance | 160 km ⁻² | 10 |
| Detailed geological mapping | 600 km ⁻² | 1 |
| Geochemical surveying orientation | 15 km ⁻² | 50 |
| Drainage survey | 50 km ⁻² | 25 |
| Soil or biogeochemical survey | 750 km ⁻² | 2 |
| <i>Geophysical surveying</i> | | |
| Resistivity and spontaneous potential | 160 km ⁻¹ | 10 |
| Induced potential | 5000 km ⁻¹ | 0.5 |
| Diamond drill cores | 40 m ⁻¹ | - |
| Shaft sinking | 5000 m ⁻¹ | - |

Exploration is no longer a haphazard process based on a prospector's intuition, individual experience and luck, although all three still play a role. As easily found resources become more scarce and economic viability of

production subject to increasing stringency, the search requires a more scientific basis. Targeting or play generation express this ethos. Explorationists base a geological model on the fund of knowledge about regional and local

structural controls on known mineralization, hydrocarbon or other resource accumulations. What is known from research about the processes involved in their formation and the blend of geological features commonly associated with such processes adds to the model. Targeting also helps define the physical and chemical peculiarities of different resources. These indicate the specific kinds of information that are needed by the explorationist to narrow down the search and eventually pin-point a deposit. This blend of critical information, known or to be gathered, forms the basis for planning and interrogating a GIS.

9.3.1 Metal exploration

Deposits of metalliferous ores form by a host of processes within the igneous and sedimentary parts of the rock cycle. There are few, if any, examples of ores originating through exclusively metamorphic processes. A large proportion of deposits show evidence which suggests that metals concentrate to ore grade through the agency of watery fluids moving beneath the surface, as in pegmatite, porphyry, contact-metasomatic, hydrothermal and secondarily enriched deposits. Such fluids commonly carry hydrogen ions, and have low pH. Hydrogen ions are the main agent that breaks the covalent bonds in many silicates and induces them to change to more stable compounds. This induces the alteration in ordinary rocks that is commonly associated with these deposit types. Moreover, the freedom for electrons to move in watery fluids means that they have a wide variety of redox potentials (Eh), which play a major role, together with pH variations, in oxidation and reduction, and the solution and precipitation of many compounds, including ore minerals. This broad group of mineralizing processes then induces considerable changes in the common rocks with which the ores are sometimes associated.

Ore formation that largely excluded the influence of subsurface watery fluids, such as magmatic segregation and placer deposits, takes place during the formation of the igneous or sedimentary hosts to the ore. Consequently, the primary process leaves products that are part and parcel of their host. This renders them difficult, if not impossible to discern on remotely sensed images, because they are orders of magnitude smaller than common features in the host. In their unaltered state, such ores often blend imperceptibly into the background. Occasionally, spectral features peculiar to the ore minerals, as with those of chromite, make it possible theoretically to highlight them using image processing. This, however, is feasible only if the ore bodies are comparable in size with, or larger than, the spatial resolution of the sensor. Most commercial chromite ores are tabular and rarely more than a few metres in thickness, and few have been discovered using remote sensing alone.

However, the geochemical affinities of metals commonly sought in magmatic segregations, such as titanium, chromium, nickel and platinum-group metals, means that it is easily possible to reduce the area of search to likely host rocks. All four cases involve basic to ultrabasic igneous rocks. In lightly vegetated terrain their low albedos readily distinguish them on images from other igneous lithologies, and from sedimentary and metamorphic rocks. In humid areas rocks rich in ferromagnesian minerals produce fertile soils rich in montmorillonite and illite clays that retain both water and nutrients, and carry different plant associations and densities.

Placer deposits are even more intractable to remote sensing, for they are often buried beneath ordinary sediment by virtue of their content of high-density minerals. However, experience suggests many likely sites for placer accumulation in drainage, shoreline and aeolian environments. Careful analysis of geomorphological features from imagery thus helps in locating potential targets.

Remote sensing becomes a powerful exploration tool in its own right when the primary and secondary processes of mineralization result in the formation of spectral anomalies. This is especially so when the anomalies extend well beyond the extent of the ore itself. Deposits formed or modified through the influence of migrating watery fluids frequently exhibit such alteration haloes spread through large volumes of rock because of pervasive fluid migration. These quite literally can appear as targets, hopefully with a paying bull's eye at the centre.

The most common style of alteration is the breakdown of feldspars and ferromagnesian minerals to a variety of clays and other hydroxyl-bearing minerals, where the SWIR range of their spectra exhibits absorptions with patterns that are unique to each species (Fig. 1.9). The general presence of hydroxylated minerals in lightly vegetated terrain (but little potential for their identification) is provided by the contrast between Landsat TM bands 7 and 5, usually expressed by a ratio image. The three SWIR bands of JERS-1 and ASTER provide more opportunity for discrimination, but the most useful data are those from hyperspectral imaging radiometers, such as AVIRIS.

A considerable number of ore deposits contain abundant sulphide minerals, of which the most common is pyrite (FeS_2), if their deposition was associated with the availability of sulphide ions and reducing conditions. Oxidizing conditions, associated with weathering and in migrating groundwater, render iron-rich sulphides unstable. They break down to sulphuric acid and a number of ferric hydroxides and complex sulphates, which are both strongly coloured and possess crystal-field absorptions in the VNIR (Fig. 1.7). As well as gossans formed over secondarily enriched sulphide deposits, simple weathering imparts these spectral features to soils. Moreover, the ferric compounds are often in the form of

colloids that move in suspension to affect broad areas. Such anomalies are characterized by higher red reflectance than in green or blue, and can be identified using ratios of Landsat TM bands 3, 2 and 1. The crystal-field absorptions in the VNIR express themselves in ratios of TM bands 4 : 3 and 5 : 4, although the positioning of Landsat MSS bands 4 and 3 is more appropriate for these features. It is worth noting that the acid conditions which sulphide weathering creates increases the abundance of hydrogen ions, so weathered sulphide ores accelerate the breakdown of silicates to clays.

Occasionally the geochemistry of hydrothermal fluids, particularly those that are highly alkaline, induces the solution and reprecipitation of silica around ore deposits. Although quartz has no unique spectral features in the reflected region, cryptocrystalline or opaline silica contains abundant water. As well as the high albedo of siliceous surfaces, cherty or opaline alteration may exhibit absorption in TM band 5 on the flank of the 1.4 μm water feature (Fig. 1.8).

This suite of alteration features has proved extremely useful in targeting a variety of hydrothermal deposits, notably supergene-enriched, porphyry and hot-spring epithermal types. Plate 6.4 shows dramatically how the combination of all three alteration types near a porphyry copper deposit express themselves, after using carefully selected combinations of both reflected and thermally emitted data. Plate 9.5 uses three Landsat TM band ratios selected to express ferric minerals, clays and albedo to highlight possible zones of hydrothermal alteration in a completely unknown area.

The effect of fluid migration on rocks is not restricted to alteration and oxidation. If conditions are reducing to slightly oxidizing and very acid then iron in its ferrous state and a number of other transition metals are soluble. They can be leached from rocks through which the fluids pass producing bleached zones. This commonly is associated with the formation of secondarily enriched sedimentary uranium and vanadium deposits. Formerly iron-rich sandstones, either grey because of finely divided pyrite or red to orange from their contained ferric minerals, become white. The anomalous colour contrast thereby allows efficient detection of areas affected by potentially mineralizing fluids of this kind.

The presence of vegetation does not necessarily preclude a search for mineralization using images alone. The high concentrations of metals, the peculiar soil Eh and pH associated with many ore deposits, and the leaching of nutrients as a result of migration of acid fluids from sulphide oxidation often suppresses the proper development of common plants. They become stunted, discoloured or simply fail to grow, thereby forming barren areas. Such geobotanical anomalies form yet another focus for image processing in target generation. Another type of anomaly encountered occasionally is where com-

mon plants burgeon as a result of extremely favourable levels of nutrients. This particularly is the case with sedimentary phosphate deposits and carbonatites containing abundant apatite ($\text{Ca}_5(\text{PO}_4)_3$).

Potentially useful as remote sensing is, there are few ore deposits that owe their discovery to the use of images alone. In fact much of the literature on remote sensing in metal exploration focuses on features associated with well-known deposits. Exploration more generally uses specific mineralization models, such as that outlined for porphyry-style deposits in Section 8.3.2. These fuse many kinds of data that have a bearing on the model, either in the form of complex image products or within a geographical information system. The products of a GIS are so numerous and often have little immediate meaning outside the exploration team that they are not illustrated here. Instead Plate 9.6 gives an example of graphic combination of various data sets for visual interpretation.

The area is part of the English Lake District, with its physiography shown by Plate 9.6(d). It is clearly dominated by dense vegetation. The generalized geology emerges from Plate 9.6(a), and the gravity and magnetic anomaly patterns are shown in Plate 9.6(b). The exploration model is orientated to hydrothermal tin, zinc and copper mineralization associated with the granite intrusions. Plate 9.6(c) is a colour image combining stream-sediment Sn, Cu and Zn analyses as RGB. Where all three elements have high values is signified by white and cream colours—these are areas of potential interest. Experience suggests, however, that significant mineralization of this type is closely associated with granite contacts, particularly in the roof zone. In Plate 9.6(e,f) the focus shifts to the vicinity of the geochemical anomaly near the granite at left of centre on the regional images, where there is also a prominent magnetic anomaly and a negative gravity anomaly that extends beyond the outcrop of the granite.

Plate 9.6(e) is an RGB image using Landsat TM data as intensity, a constant saturation and gridded Sn values as hue, with magnetic contours overlaid. The magnetic high probably relates to magnetite formed during contact metamorphism at the roof of the buried granite. The high Sn values are spatially related to two indistinct linear features—possibly fracture zones related to deep-seated faults—trending WSW–ENE and NW–SE. Plate 9.6(f) supplements this by combining directionally filtered TM data as intensity, gravity anomaly as hue and significant Sn and Bi anomalies as red and green overlays. The purple area indicates where the granite is lurking at shallow depths. The conclusion is that tin anomalies here are spatially related to the buried granite and arranged along deep-seated fractures, and so constitute an ideal target for detailed follow-up in this poorly exposed area.

9.3.2 Hydrocarbon exploration

Oil and natural gas fields form much larger targets than metal ore deposits, but for obvious reasons they are always buried and therefore hidden. Locating them has to be precise, and relies on detailed interpretation of seismic reflection data and, at the end of the day, a commitment to the drill. The focus on a particular area on which to commit costly seismic surveys and drilling relies almost entirely on play generation. A hydrocarbon play is a combination of favourable geological circumstances for the likely presence of oil or gas, conditioned by the economics involved in developing any finds in a particular region. The geological aspects of a play can be summarized as the coincidence of a number of favourable factors: the presence of an organic-rich source rock; its burial to a depth where local geothermal conditions permit the maturation of organic compounds to oil and gas; a path whereby migration from source to reservoir can occur; the presence of rocks with sufficient porosity and permeability to act as a reservoir; the presence of a suitable trap with an impermeable *cap* of some kind where the hydrocarbons can accumulate and be preserved. Previous drilling records in partly developed oil fields, together with stratigraphic and structural interpretation of seismic sections and exposed geology permit the assessment of many of these factors. The huge amount and diversity of data in well-known hydrocarbon basins, including those from existing fields, are now handled and analysed routinely in what amount to a three-dimensional GIS. Remote sensing methods sometimes complement such analyses, but assume a more important role in unexplored frontier areas. That is the main focus here.

All significant hydrocarbon fields are found in or near sedimentary basins that subsided sufficiently to permit the deposition of thick and varied sedimentary sequences. There are many different tectonic situations where these basins might have formed, and most large basins are already known. Without expensive seismic surveys, however, the details of their architecture are sketchy. Thick accumulations of sediments with densities and magnetic susceptibilities lower than crystalline basement give rise to negative gravity anomalies and magnetically 'smooth' zones. So a first approximation of their innards arises from examination of gravity and magnetic data. Using gridded data and derivatives from them in image form, as in Figs 8.11–8.15, helps locate variations in basin depth. Knowledge of subsidence rates and geothermal conditions helps assess zones within the oil and gas maturation 'windows'. Sometimes, gravity and magnetic data help visualize deep fracture zones that may have focused sedimentary facies variations and fluid migration. The radar altimetry of the oceans discussed in Section 8.2.4 (Fig. 8.10) contains information on regional variations in

gravitational potential. So actual gravity measurements are not always essential. Many offshore basins show up on derivatives of radar altimetry.

Analysis of surface exposures gives clues to the internal structure of sedimentary basins. Chapters 4, 6 and 7 show many examples of the structural and lithological information that can be won directly from remotely sensed images. Although this is unlikely to locate oil fields directly, it provides a basis for basin analysis and understanding of the tectonic controls over basin evolution. Plates 4.3, 4.4, 6.5 and 7.1, and Figs 7.5, 7.8 and 7.10 are particularly instructive in this respect.

Natural oil and gas seepages, sometimes flaring due to ignition by lightning, were the primary indicators of a number of oil fields, both onshore and at sea. Every oil explorationist hopes to make such an 'easy' discovery, for little remains other than to investigate by drilling. There is no such thing as a completely 'tight' trap, particularly with respect to light hydrocarbons and gas. Every hydrocarbon field undergoes some leakage, or has leaked in the past. Whether or not this reaches the surface, and has a sensible effect if it does, is a matter of chance. Remote sensing for seep detection is consequently a topic in which considerable resources have been invested. It must be emphasized that evidence for seepage is rarely a signpost to 'drill here'. Light petroleum and gas migrate along the lines of greatest permeability, either a suitable sedimentary stratum or a zones of fracturing. Sometimes these might be directly above a trap, but equally they may be orientated laterally. Also, the signs may indicate past leakage, that has exhausted the source which may now contain only highly viscous tars. A good example of the latter are tar sands, from which the light fraction has escaped, and which require mining and heat treatment to win useful products.

The most obvious seepages contain liquid oil and tarry residues from evaporation, and everyone knows that tar is black. Good examples are the La Brea tar ponds of the Ventura basin in Los Angeles, and the major Iranian fields were found from flaring seeps known for centuries by local people. Dark signatures on images from the reflected region in prospective basins are always worth noting. However, most seeps are far more subtle. There have been reports of a sort of physical blurring of minor physiographic features above several major, on-shore oil fields in unvegetated terrains. This has not yet been explained, but may be due to periodic upwelling of formation water charged with gas to create quicksand conditions. Of more use are the potential effects of such fluids on the chemical and physical properties of the materials through which they have passed.

Accumulations of hydrocarbons require extremely reducing conditions, and the formation waters associated with them have low Eh. The passage of reducing fluids through rocks containing oxidized iron minerals, such as

hematite, induces the reduction of insoluble ferric iron (Fe^{3+}) to ferrous (Fe^{2+}) ions, which are soluble. One possible effect of escaping hydrocarbons and the movement of formation water is bleaching, particularly of red beds in the sequence. Not only iron can be mobilized in this way, but many other metals too. So, seepage zones can be associated with anomalous soil geochemistry. Geochemical soil surveys, for hydrocarbons as well as metals, can figure in hydrocarbon exploration, where the methods described in Section 8.2.3 are appropriate.

One of the most controversial methods in hydrocarbon exploration centres on another likely phenomenon in highly reducing environments. Low Eh can partly reduce the common ferric oxide and hydroxide colourants in many sediments to black magnetite (Fe_3O_4). Little imagination is required to guess that seepages associated with such diagenetic magnetite may induce high-frequency magnetic anomalies. The controversy lies in whether it happens at all, and if it does whether the anomalies can be screened from others with different sources. In active exploration areas the ground will be cluttered with all sorts of ironmongery that would produce very similar anomalies. That it does occur is partly confirmed by increased magnetic susceptibility in cores from productive wells to depths as great as 1 km compared with that in dry holes. This has been backed up by the discovery of tiny magnetite spherules in the cores. Their form encourages the idea that as well as reducing conditions some kind of biological activity is involved, of which more shortly. Conceivably, the diagenetic magnetite phenomenon adds more value to the use of high-resolution aeromagnetic surveys in oil exploration, in which sophisticated spatial-frequency filtering and image enhancement may isolate seepage-related anomalies. As unconsolidated sediments on the sea floor have more potential for the passage of seeping fluids, and offshore exploration is dominated by high-cost seismic surveys, related methods, including remote measurements of magnetic susceptibility of the sea bed, are under development.

Far less controversial are methods related to the effects of noxious seeping fluids on terrestrial plant life. Plants obtain their nutrients and water through complex symbioses of bacteria and fungi at their root tips. If these are damaged then the plants die, fail to thrive or show unusual features. Enhanced hydrocarbon content in groundwater encourages the growth of anaerobic bacteria that reduce sulphate (SO_4^{2-}) to sulphide (S^{2-}) ions. This is the source of hydrogen sulphide or 'sour gas' in many oil fields, and its presence encourages the precipitation of heavy-metal sulphides carried by groundwater or seeping formation water. Sour-gas generation is the reason for the association of geochemical anomalies with some seeps, and is possibly involved in diagenetic magnetite formation. It and the 'dumping' of heavy metals wreaks havoc on root-tip symbioses, with a range of effects

on overlying plant life. They may become stunted or discoloured, wholesale die-off may occur, and the timing of seasonal colour change in the canopy (Section 1.3.3, Fig. 1.17) may change, commonly resulting in premature senescence (Plate 9.7). In vegetated areas remote sensing provides a wide range of opportunities to locate such seep-related geobotanical anomalies.

Hydrocarbon seepage is not restricted to land areas, but it would be easy to dismiss as a focus for offshore exploration because the sea floor is not usually accessible to remote sensing. But consider for a moment what gas seepage beneath water might do. It forms rising plumes of bubbles. Bubbles do not rise passively, but induce a slow upward component of movement in the water. This is the principle of the air-lift dredge used in marine archaeological excavations. Such plumes can draw cold, nutrient-rich bottom water to the sea surface. Of course, other phenomena such as tidal and current activity result in upwellings, but these change position with the seasons. A seep-related upwelling should remain in the same place. Theoretically therefore images of sea-surface temperature and colour associated with phytoplankton blooms, taken over a period, should reveal fixed cold spots and planktonic blooms that are possibly related to seepage. As offshore interest is in large hydrocarbon fields with commensurately large seeps, such an indirect exploration method can be conducted on a daily basis using low-cost meteorological data with coarse resolution. A number of major oil companies have conducted research along such lines under conditions of great secrecy. Related to it is the deployment of low-level airborne surveys of aerosols at the sea surface that result from the bursting of small bubbles. The hydrocarbon chemistry of the sucked-up samples is monitored using gas chromatography. However, growing knowledge that many areas of the sea floor are underlain at shallow depths by sediments saturated with methane hydrates, formed by reaction between decay products from plankton and cold, high-pressure water, means that many submarine seepages are not related to commercial oilfields.

9.3.3 Water exploration

In the midst of a world economic depression, many of the talents of remote-sensing geologists are being shed by oil and mining companies, as demand and price of their products stagnate or fall. At the same time, in both developed and underdeveloped regions, the problems of water shortage and decline in its quality grow to disastrous proportions. A sage once commented that the first provision in any civilized society, after a means of collecting taxes, is that of a safe and dependable water supply, yet this basic need is under considerable threat. To a large extent geologists can assemble the information needed to alleviate, if not solve these problems.

Water supply is served by two main options, outside of desalination and other technological fixes. These centre on surface water and that stored in rocks and soils as groundwater. The use of remote sensing in hydrology and hydrogeology relates mainly to suggesting favourable areas for development in terms of the relative amounts of water available and various factors that might condition development. Indirectly, it has some bearing on water quality. An example would be the identification of zones of hydrothermal alteration and beds of marine shales, both of which can contribute potentially toxic compounds to waters flowing over or through them. Screening for biological pollution would in most cases be served through a GIS by data layers on habitations and land use. A geologically or hydrologically favourable site may not necessarily be useful. To evaluate it in economic and social terms requires information on the type and location of usage, as well as on various engineering considerations. Again, a multifaceted GIS is required, of which geological information from remote sensing and other spatial data would form a part.

The surface water regime in an area can be expressed by the relationship between the four factors of precipitation, surface flow, evapotranspiration and infiltration. Only when channelled or in lakes does surface water become a potential resource. Aside from losses by evapotranspiration and infiltration, the flow through a channel or entering a lake depends on the area of upstream catchment together with precipitation. The suitability of the water, in terms of sediment load, depends mainly on the slope angles in the catchment and the density of vegetation cover that binds soil.

Precipitation information can be derived from the use of monthly averages of cloud-top temperature on meteorological satellite (metsat) images as a gauge of the likelihood of rain falling from any clouds, and by the use of multi-temporal images showing snow cover and snow melt rates. Evapotranspiration rates can be estimated from surface temperature averages derived from metsats, and by models based on vegetation density derived from its unique contrast in VNIR and red reflectance. Infiltration is a thornier problem and is discussed later. Catchment area and slope estimation, together with analysis of drainage networks, are derivatives from DEMs, as covered by Section 8.2.4. Combined with remotely sensed estimates of vegetation cover, this allows the qualitative assessment of each catchment for likely levels of turbidity resulting from soil erosion.

The main input by geologists to studies of surface water resources focuses on the location and evaluation of sites for capturing flow and harvesting it for drinking water, irrigation or hydropower generation. Whereas long earth barrages or *bunds* pond water in broad valleys on gently undulating terrain, the best sites are where valleys narrow as they pass through more resistant

rocks. These thresholds are often floored by thin mantles of permeable alluvium so providing firm footings for the construction of dams or barrages. Thresholds are easily located on stereoscopic aerial photographs or satellite images (Fig. 4.13). Upstream erosion always poses the problem of siltation of artificial reservoirs, the useful lifetime of which may be reduced to only a few years in areas of high soil erosion. Although this might seem to suggest that surface dams are best avoided in such areas, there is an interesting means of turning siltation to advantage. Provided the sediment fill behind a dam has good porosity and permeability, once completely silted it can still retain useful water supplies, accessed by infiltration galleries built deliberately to be buried by sedimentation. Such artificial groundwater reserves have the advantage over open water reservoirs of being less prone to loss by evaporation in hot areas. The main reason why transported sediments are unsuitable for this secondary storage potential is the filling of their pore spaces by fine grains or clays. If the supply of sediment is free of fines and stream velocities are energetic enough to move sand and gravel, dams designed to fill and produce from their eventual sedimentary fill are a promising option. The best sediments are those composed mainly of quartz particles. Remote sensing helps evaluate potentially useful catchments for this new strategy in terms of the bedrock exposed to erosion. Areas of granite or older sandstones should source sediments with good hydraulic properties. Rocks that easily rot to clays, such as basic igneous rocks, mudstones and fine-grained metamorphic rocks produce sediments with no secondary supply potential.

Hydrogeological assessment of groundwater resources involves primarily the mapping of lithologies and structures to help assess the subsurface disposition of sedimentary aquifers. This is akin to basin analysis for hydrocarbons, with the difference that groundwater migrates laterally and downwards. The conditions for subsurface water accumulation, involving various kinds of trap, are familiar to most geologists. Images provide abundant information that can be blended with observations on runoff versus infiltration potential in recharge areas. Here the intention is to concentrate on specific advantages offered by remote sensing.

Among the many sites for groundwater accumulation are those provided by subsurface dams. These include faults that juxtapose aquifers with aquitards, and igneous dykes with a crystalline nature that blocks flow where the dyke trends across the direction of subsurface flow. Both faults and dykes exhibit discontinuous linear features at the surface that often are more easily seen and extrapolated on images (Figs 4.12, 4.24 and 7.9, and Plates 4.2 and 4.3) than on the ground. Radar is particularly useful in locating small dykes because it highlights the aligned corner reflectors formed by jointing (Fig. 7.9),

even when the dykes are buried by veneers of alluvium in arid areas (Fig. 7.17). Because dykes are frequently more resistant than their hosts, when a landscape becomes mantled by superficial soft sediments older dykes may form corrugations at the interface, which pond groundwater or slow its flow.

Unconsolidated sediments, especially piedmont fans, alluvium, aeolian sand and beach deposits (Figs 4.42, 4.43, 4.45 and 4.51), provide limited supplies of groundwater that are easily accessible. In semiarid climates the water table in such sediments may be deeper than the extent of the potential host, so estimates of thickness are needed. Apart from geophysical surveys at the surface, the only means of visualizing this parameter is the use of radar images. Provided soil moisture is low, radar can penetrate as deep as 4 or 5 m to give a signal from subsurface features (Section 7.1.1). Figure 7.17(a) shows a dramatic example from the Red Sea Hills of Sudan, where otherwise completely hidden intricacies of a drainage system buried by alluvium indicate the most likely sites for obtaining water from the main distributaries of the old drainage. Conversely, in areas of variable soil moisture—a good guide to deeper groundwater—radar scatterometry gives a guide to moisture content.

Whereas faults in sedimentary sequences provide opportunities for subsurface dams, in areas dominated by impermeable crystalline rocks the shattering associated with faulting provides the only opportunity for infiltration and eventual recovery of groundwater. Remote sensing is pre-eminent in delineating such zones. Of special interest is the extrapolation of large fracture systems from elevated areas of high precipitation to lower, drier ground. The fractures potentially can transfer water laterally over tens to hundreds of kilometres. Because the secondary porosity involved in fracturing can become sealed by fines, evaluating fracture zones requires consideration of the rock types involved and the nature of weathering. In all climates micaceous rocks are easily sealed. In humid areas, or those once subject to high chemical weathering, feldspars in granitic and gneissic rocks, and ferromagnesian minerals in more mafic rocks break down to clays that clog the potential pathways. In dry climates many granites break down by mechanical means to form gravely regoliths and fracture zones still retain their permeability. In all climates, inert quartzites and soluble carbonates are the ideal hosts for fracture-controlled aquifers. As discussed in Section 4.2, most images carry clues to the general distribution of lithologies, and multispectral data can help further distinguish rock types.

Because phreatophyte plants can put down roots to depths in excess of 100 m to reach the water table, the presence of vegetation in arid areas is a sure guide to the presence of water, although not necessarily to its quality. Plate 9.8 shows an example of the discontinuous

distribution of vegetation in an arid, faulted area, and clearly indicates areas favourable for development of subsurface supplies. Phreatophytes in fact consume large volumes of groundwater. One estimate is that one square kilometre of dense desert scrub transpires sufficient water vapour in a year to supply all the needs of a village of 1000 inhabitants.

9.4 Engineering applications

Geologists have long made important contributions to the siting of major constructions, such as roads, dams, power plants and waste disposal sites. Much of this has centred on aspects of safety and stability in relation to various kinds of slope failure, seismicity and flooding, as well as to the engineering properties of foundations and sources of potential construction materials.

Slope stability and risk from landslips are most appropriate to routing of roads and railways. Although there is a variety of contributory factors to instability, the common factor to all such hazards is variation in topographic slope. Stereoscopic images help us visualize slope, but these generally incorporate a degree of vertical exaggeration. It requires painstaking use of photogrammetric instruments to produce accurate estimates of slopes and slope profiles. A DEM makes it a simple matter to plot topographic profiles, to extract slope values and to highlight those so steep as to pose a hazard of failure.

In humid tropical areas, where thick soil and regolith potentially pose problems of mudflows and arcuate slips, extremely steep slopes are commonly stabilized by dense vegetation and the associated root mat. Vegetation cover, or the lack of it, is an important parameter in slope-stability assessment in soil-covered areas. Together with gradient data, multispectral means of vegetation classification from images clearly forms an important input to planning. In areas largely free of soil, vegetation is likely to give little assistance to the engineer. The rock type and associated structures govern the likelihood of rock falls and landslips. The most risky combination is that of interbedded permeable and impermeable strata which dip parallel to the topographic slope. Both gravity and lubrication by water along bedding surfaces conspire to encourage detachment and mass wasting. Generally, satellite imagery has too coarse resolution to permit detection of such features, although multispectral data may assist in delimiting lithologies with contrasted engineering properties. Stereoscopy using aerial photographs is a decided advantage in locating potential hazards and, equally important, evidence of previous slope failure.

Steep slopes are not the only areas posing potential hazard to heavy construction. Soft sediments, such as peats and clay-rich rocks, are easily compacted as water is expelled during loading. If the compaction varies

laterally, the differential subsidence can cause failure. Former lake beds, glacial tills and flood-plain alluvium are replete with such hazards, yet they are often signified on images by the distinctive landforms associated with their formation (Figs 4.43, 4.47 and 4.50) that are not easily detectable on the ground.

Section 4.4.1 illustrated several of the features associated with active faults, most of which require the fine resolution of aerial surveys. Now it would be a pretty dim geologist who failed to recognize the hazards associated with a project entering seismically active terrain, because virtually all sizeable earthquakes above magnitude 4 are readily and systematically located by a global network of seismographs. The crucial issue is to relate them to fault zones, although destructive earthquakes along subduction zones have no surface manifestation, having deep epicentres. A first call would be to delineate all potential faults that might fail in a seismically risky area, using aerial photographs. Most damage in seismically active areas, however, occurs not in high-relief areas of hard rocks, where faults are easily spotted, but in areas of unconsolidated sediment that behave plastically during earthquakes, even when the seismic event is far distant. The catastrophes of Mexico City (1988) and Turkey (1999) bear grim witness to this. Areas of silts and clays, often associated with former lake beds and estuarine deposition, need to be delimited. This is no easy task, as all low-lying flat areas are suspect. Careful searches for morphological features, such as strand lines (Fig. 4.51), that indicate such sediments must be conducted.

Equally, it would be an engineer verging on the psychotic who wilfully built a bridge, a dam or a power plant straddling a fault of any age, for the loading itself could induce movement. These criminal errors do occur, however. Old lines of weakness are often exploited by modern earth movements, such as those related to post-glacial unloading. Displacements of centimetres could cause catastrophic events in structures containing moving parts with low tolerances, such as the control rods in a nuclear plant. However, it is not always easy to recognize old fractures on the ground or even on images, because their active control over landforms is long extinct. In areas of low relief and intensive agriculture the patchwork of fields with different crops produces an efficient camouflage for muted topographic features that might indicate faults (Fig. 4.49). Various image processing methods are available to reduce this cultural camouflage and reveal previously unsuspected features of tectonics (Fig. 9.5). Radar interferometry helps express the kinds of small-scale ground movement associated with earthquakes (Plate 9.9d). Although this is of no use in mitigating effects of any particular earthquake, by showing the breadth subject to ground movements, research using InSAR might serve to indicate corridors of risk along known fault lines.



Fig. 9.5 Enhanced early spring Landsat TM band 5 image of part of an intensely farmed area in southern England. At this time of year grasses and winter cereals have similar moisture contents and the field patterns that generally obscure topographic features (Fig. 4.48) are uniquely removed to reveal many major fault zones that are unreported by conventional geological mapping.

Whether they move or not, fractures form pathways for the large-scale movement of fluids (Sections 9.3.2 and 9.3.3). This is the last thing that is desired in a waste-disposal site, particularly one for long-lived nuclear debris. Besides using images to screen areas with faults from consideration, many of the techniques deployed for water exploration are useful in selecting waste-disposal sites. Apart from methane generation produced by fermentation of organic waste that vents upwards and laterally, leakage follows exactly the same rules as the movement of groundwater, and hydrogeological interpretation of image data is essential in siting.

One hazard that engages the undivided attention of insurance actuaries in urban areas is that of ground subsidence, which can cause minor though expensive damage to buildings and services. It occurs most frequently in two main ways, through subsurface excavations, such as coal mining or solution mining of salt, or because of the volume reduction of aquifers when water extraction exceeds rates of groundwater recharge. Movements

are rarely more than a few centimetres, although their cumulative effects can build to metres. Consequently, they are difficult to detect remotely, although the damage is all too easily witnessed. The ability of interferometric radar imaging to create both highly precise DEMs and images of elevation change opens up an entirely new field of subsidence detection and policing, for the victims are otherwise very rarely able to prove who, if anyone, is responsible. Plate 9.9(a,b) shows InSAR results for subsidence related to both underground mining and groundwater abstraction.

Although not strictly speaking an issue for which there are engineering remedies, volcanism and the associated hazards of debris flows form an extreme risk in volcanically active regions. There is absolutely no problem in recognizing volcanoes, but the most hazardous are those that remain dormant for long periods between major eruptions. Soils formed from lavas and ash are among the most fertile there are, and farmers are drawn to till them despite the risk, particularly where good land is at a premium. Communities grow, and in the Andes some of the most hazardous areas on the planet support high populations on the lower slopes of volcanoes. Eruptions of lavas pose little hazard to life, simply because they are quite obvious, and travel relatively slowly, although flows will destroy everything in their path. Having temperatures in excess of 1000°C, active flows and lava lakes emit radiation at wavelengths as short as the visible range. They closely approximate black bodies, and the radiance in several wavebands allows temperature estimation using the Stefan-Boltzmann Law. Detecting eruptions by remote sensing is easy, and although seemingly altruistic is akin to telling the grandmothers on the slopes of a volcano how to suck eggs. It is an excuse for academic study of volcanism—perfectly justifiable but an excuse nonetheless. Detecting signs that a volcano is about to 'go off' is altogether more urgent, for two main reasons. First, high-altitude dormant cones are often capped by snow and ice, which if melted—even at their base—may generate avalanches and huge mudflows known as *lahars*. These travel at speeds far in excess of lavas and account for the majority of fatalities. The second are even more rapid incandescent debris flows, formed when volcanoes being inflated by viscous magma fail sideways instead of vertically through a vent to produce an ash plume. Figure 4.41(b) shows features typical of such a catastrophic event, as do the film sequences of the fate of Mount St Helens in May 1980. The ruins of Pompeii and Herculaneum near Vesuvius bear witness to the awful consequences for anyone unfortunate enough to live close to volcanoes with such tendencies.

Signs of pending eruption, of whatever kind, are subtle. There may be an increase in fumarole activity, perhaps heating at the site of a future eruption. Neither of these is easily monitored by remote sensing, partly

because they are minor features, often smaller than a system's resolution, and partly because their signatures are easily overlooked. A large body of detailed evidence suggests that the slow rise of magma before a volcanic event induces slight changes in both the gravitational field around volcanoes and their topography, as well as producing swarms of minor earth tremors. Detecting these and other signs by continual field studies is in its infancy and focused on famous volcanoes such as Etna. Interferometric radar provides a means of precisely documenting volcano shape changes (Plate 9.9c) that potentially might become a means of early warning. The ERS-1 and -2 orbiting radar imaging systems are ideal for monitoring of that kind, but such is the complexity of InSAR processing that few institutions yet have the capacity for keeping watch over one volcano, let alone all those that pose real threats.

9.5 Geochemical hazards

Statistical analysis of medical records, particularly those of morbidity and mortality (sickness and cause of death), are beginning to reveal disturbing evidence for health blackspots that correlate with a variety of environmental parameters. These are in addition to the known associations of mental retardation with lead pollution, leukaemia and other cancers with nuclear plants, a variety of horrible afflictions with mercury and cadmium pollution and birth defects clustered around sites for hazardous waste disposal, and seem to stem from purely natural environmental variations.

Some environmental hazards have been known for a long time, such as the associations of goitre and fluorosis with areas deficient in iodine and with abnormally high fluorine contents in soil, respectively. However, new relationships are appearing regularly. Some cancers are more frequent over certain uranium-rich granites and shales, and have been traced to radon-gas emissions from the soil accumulating in houses. There is some evidence for increased incidence of multiple sclerosis in fluorine-rich areas. Men living in areas supplied with water from limestones are more at risk from heart disease than those in soft-water areas. High contents of organic pigment from peat in water correlates with increased risk of stomach cancer. To these few examples of human risk from geochemical anomalies can be added incidence of ill thrift in livestock from excess molybdenum, copper deficiency and lead anomalies. Whereas the associations are known, currently the geochemical data to assess variable risk are lacking for most parts of the world.

The most useful information would be results from geochemical surveys of soil, as carried out routinely but in restricted areas for metal exploration. This is extremely expensive (Table 9.1), and unlikely ever to be

conducted globally. At much lower cost, and available now for many developed countries, are data from drainage surveys (Section 8.2.3). In image form (Plate 9.6) they provide an extremely efficient means of screening for geochemical hazards, as they are for exploration, and could be incorporated easily into health-orientated GISs to be assessed with the many other variables. Risk from naturally high radioactivity is readily assessed by reference to images of gamma-ray emissions derived from airborne radiometry (Plate 8.4).

In the absence of analytical data, the known concentration or depletion of some of the geochemical culprits with specific rock types and with relatively unusual processes, such as those involved in mineralization, means that geological maps are a rough guide to potential problems. Maps derived from images are in some cases even more useful in this regard than those compiled by more conventional means. The discussion in Section 9.3.1 on the signatures of hydrothermally altered areas and those where sulphides have been oxidized to products including sulphuric acid shows that remote sensing has a unique contribution in this respect. Hydrothermal and other types of mineralization involving pervasive fluids boost the levels of many hazardous elements in rocks and their associated soils. The natural release of acids takes some of these elements into solution, further dispersing them in the natural environment.

There is one important point of an ethical nature, which is often overlooked in assessment of hazards of whatever kind. It may well be possible to establish a previously unsuspected risk. The problem lies in informing the people exposed to it. If there is no remedy, other than migration at great cost and carrying perhaps yet other risks, should they be told?

A final point of ethical content is perhaps obvious, but rarely stated. As well as intelligence analysts, industrial and academic geologists in the developed countries are privy, through remote sensing and digital image processing, to a great deal of strategic economic information about less-fortunate areas of the world. Often they are able to learn more about resources and terrain than their counterparts in the so-called 'Third World' because of their technological advantage. In an ideal world, such information should be transmitted on a completely equitable basis, but frequently it is not. The simplest recourse is the spread of the enabling technology of remote sensing, and costs of hardware and software to achieve what is outlined in this book have now been reduced to levels that are affordable by any government. The question of affordability of data, for a continuous supply is needed to put theory into practice, is a different matter. Whereas Landsat-7 is back in the domain of public access at low cost, and data from the EOS programme will be free to *bona-fide* researchers, most data, both existing and planned,

remain in the commercial domain at a price beyond the purses of most of those who would benefit from them. It remains to be seen whether the proliferation of orbiting systems will flood the market and force down prices, or if a kind of cartel among the operating companies will maintain them artificially.

Further reading

These references are but a few of the many relating to applications of remote sensing and GIS in geology. The *Proceedings of the Thematic Conferences on Remote Sensing for Exploration Geology*, sponsored by the Environmental Research Institute of Michigan (ERIM), include many papers in this very general area. Occasionally, trade journals, such as the *Mining Magazine* and the *Oil and Gas Journal* will contain review articles on this topic. Journals, such as *Economic Geology* and the *Bulletin of the American Association of Petroleum Geology*, often carry papers on remote sensing in exploration, but so do most Earth Science journals. A good way to keep abreast of papers on remote sensing in geology is using the search options in the Science Citation Index, or in the electronic journals lists of various publishers, using appropriate search terms and keywords.

- Abrams, M.J., Brown, D., Lepley, L. & Sadowski, R. (1983) Remote sensing for copper deposits in Southern Arizona. *Economic Geology* **78**, 591–604.
- Abrams, M.J., Conel, J.E. & Lang, H.R. (1985) *The Joint NASA/Geosat Test Case Study*, 2 Volumes. American Association of Petroleum Geologists, Tulsa, Oklahoma.
- Abrams, M., Bianchi, R. & Pieri, D. (1996) Revised mapping of lava flows on Mount Etna, Sicily. *Photogrammetric Engineering and Remote Sensing* **62**, 1353–1359.
- Abrams, M.J., Ashley, R.P., Rowan, L.C., Goetz, A. & Kahle, A.B. (1977) Mapping of hydrothermal alteration in the Cuprite mining district, Nevada, using aircraft scanner images for the spectral region 0.46–2.36 μm . *Geology* **5**, 713–718.
- Allum, J.A.E. (1981) Remote sensing in mineral exploration—case histories. *Geoscience Canada* **8**, 87–92.
- Alwash, M.A. & Zilger, J. (1994) Remote sensing-based geological mapping of the area west of Al Madinah, Saudi Arabia. *International Journal of Remote Sensing* **14**, 163–172.
- Bailey, G.B. & Anderson, P.D. (1982) Applications of Landsat imagery to problems of petroleum exploration in Qaidam Basin, China. *Bulletin, American Association of Petroleum Geologists* **66**, 1348–1354.
- Baker, M.C.W. & Baldwin, J.A. (1981) Application of Landsat multispectral classification for locating gosans in North Chile. In: *Remote Sensing in Geological and Terrain Studies* (eds J.A. Allan & M. Bradshaw), pp. 25–33. Remote Sensing Society, London.
- Berhe, S.M. (1988) A regional tectonic study of NE and E Africa and its implication for mineral exploration: a synoptic view from satellite imagery. *Proceedings, IGARSS '88 Symposium, Edinburgh*, pp. 363–366. ESA SP-284, European Space Agency, Paris.

- Billings, W.P. (1950) Vegetation and plant growth as affected by chemically altered rocks in the Western Great Basin. *Ecology* 30, 62–74.
- Birnie, R.W. & Francica, J.R. (1981) Remote detection of geobotanical anomalies related to porphyry copper mineralization. *Economic Geology* 76, 637–647.
- Bolviken, B., Honey, F., Levine, S.R., Lyon, R.J.P. & Prelat, A. (1977) Detection of naturally heavy-metal-poisoned areas by Landsat-1 digital data. *Journal of Geochemistry and Exploration* 8, 457–471.
- Bushara, M.N. (1996) Neogene geological structures using enhanced Thematic Mapper over eastern Zagros Mountains. *International Journal of Remote Sensing* 17, 303–313.
- Chang, S.H. & Collins, W. (1983) Confirmation of the airborne biogeophysical mineral exploration technique using laboratory methods. *Economic Geology* 78, 723–736.
- Cole, M.M. (1982) Integrated use of remote sensing imagery and geobotany in mineral exploration. *Transactions of the Geological Society of South Africa* 85, 13–28.
- Collins, W., Chang, S.H. & Kuo, J.T. (1981) Detection of Hidden Mineral Deposits by Airborne Spectral Analysis of Forest Canopies. NASA Contract NSG-5222, Final Report.
- Collins, W., Chang, S.H., Raines, G., Channey, F. & Ashley, R. (1983) Airborne biogeochemical mapping of hidden mineral deposits. *Economic Geology* 78, 737–749.
- Conradsen, K. & Harpoth, O. (1984) Use of Landsat multispectral scanner data for detection and reconnaissance mapping of iron oxide staining in mineral exploration, central East Greenland. *Economic Geology* 79, 1229–1244.
- Crosta, A.P., Sabine, C. & Tarantik, J.V. (1998) Hydrothermal alteration mapping at Bodie, California, using AVIRIS hyperspectral data. *Remote Sensing of Environment* 65, 309–319.
- Darch, J.P. & Barber, J. (1983) Multi-temporal remote sensing of a geobotanical anomaly. *Economic Geology* 78, 770–782.
- Das, J.D., Saraf, A.K. & Jain, A.K. (1995) Fault tectonics of the Shillong Plateau and adjoining regions, north-east India using remote sensing data. *International Journal of Remote Sensing* 16, 1633–1646.
- Dennis, A.M., Harris, A.J.L., Rothery, D.A., Francis, P.W. & Carlton, R.W. (1998) Satellite observations of the April 1993 eruption of Lascar volcano. *International Journal of Remote Sensing* 19, 301–321.
- Deroin, J.-P., Motti, E. & Simonin, A. (1998) A comparison of the potential for using optical and SAR data for geological mapping in an arid region: the Atar site, Western Sahara, Mauritania. *International Journal of Remote Sensing* 19, 1115–1132.
- De Souza Filho, C.R. & Drury, S.A. (1998b) Evaluation of JERS-1 (Fuyo-1) OPS and Landsat TM images for mapping of gneissic rocks in arid areas. *International Journal of Remote Sensing* 19, 3569–3594.
- Deutch, M. & Estes, J.E. (1980) Landsat detection of oil from natural seeps. *Photogrammetric Engineering and Remote Sensing* 46, 1313–1322.
- Donovan, T.J., Forgey, R.L. & Roberts, A.A. (1979) Aeromagnetic detection of diagenetic magnetite over oil fields. *Bulletin, American Association of Petroleum Geologists* 63, 245–248.
- Drury, S.A. (1990) Satellite aid. *New Scientist* 126, 50–53.
- Drury, S.A. (1990) SPOT image data as an aid to structural mapping in the southern Aravalli Hills of Rajasthan, India. *Geological Magazine* 127, 195–207.
- Drury, S.A. & Berhe, S.M. (1991) Remote sensing in groundwater exploration in arid regions: examples from the Red Sea Hills of NE Africa. *Association of Geoscientists In Development (AGID) News* 67/68, 17–20.
- Drury, S.A. & Berhe, S.M. (1993) Accretion tectonics in northern Eritrea revealed by remotely sensed imagery. *Geological Magazine* 130, 177–190.
- El Rakaiby, M.L. (1995) The use of enhanced Landsat TM image in the characterization of uraniferous granitic rocks in the Central Eastern Desert of Egypt. *International Journal of Remote Sensing* 16, 1063–1074.
- Ellyett, C.D. & Pratt, D.A. (1975) A Review of the Potential Applications of Remote Sensing Techniques to Hydrogeological Studies in Australia. Technical Paper 13, Australian Water Research Council, Canberra.
- Farrand, W.H. (1997) Identification and mapping of ferric oxide and oxyhydroxide minerals in imaging spectrometer data of Summitville, Colorado, USA, and the surrounding San Juan Mountains. *International Journal of Remote Sensing* 18, 1543–1552.
- Feder, A.F. (1986) Integrated interpretation for exploration. *Oil and Gas Journal May*, 180–184.
- Feldman, S.C. & Tarantik, J.V. (1988) Comparison of techniques for discriminating hydrothermal alteration minerals with airborne imaging spectrometer data. *Remote Sensing of Environment* 24, 67–83.
- Fookes, P.G., Sweeney, M., Manby, C.N.D. & Martin, R.P. (1985) Geological and geotechnical engineering aspects of low-cost roads in mountainous terrain. *Engineering Geology* 21, 1–152.
- Francis, P.W. & De Silva, S.L. (1989) Application of the Landsat Thematic Mapper to the identification of potentially active volcanoes in the Central Andes. *Remote Sensing of Environment* 28, 245–255.
- Francis, P.W. & Rothery, D.A. (1987) Using the Landsat Thematic Mapper to detect and monitor active volcanoes: an example from Lascar volcano, northern Chile. *Geology* 15, 614–617.
- Fraser, A., Huggins, P., Rees, J. & Cleverly, P. (1997) A satellite remote sensing technique for geological structure horizon mapping. *International Journal of Remote Sensing* 18, 1607–1616.
- Goosens, M.A. & Kroonenberg, S.B. (1994) Spectral discrimination of contact metamorphic zones and its potential for mineral exploration, Province of Salamanca, Spain. *Remote Sensing of Environment* 49, 331–344.
- Gupta, R.K. & Badarinath, K.V.S. (1993) Volcano monitoring using remote sensing data. *International Journal of Remote Sensing* 14, 2907–2918.
- Gupta, R.P. & Bodechtel, J. (1982) Geotechnical applications of Landsat image analysis of Bhakra dam reservoir, India. *Remote Sensing of Environment* 12, 3–13.
- Gupta, R.P. & Joshi, B.C. (1990) Landslide hazard zoning using the GIS approach—a case from the Ramganga catchment, Himalayas. *Engineering Geology* 28, 119–131.
- Halbouty, M.T. (1976) Application of Landsat imagery to petroleum and mineral exploration. *Bulletin, American Association of Petroleum Geologists* 60, 745–793.
- Halbouty, M.T. (1980) Geological significance of Landsat data for 15 giant oil and gas fields. *Bulletin, American Association of Petroleum Geologists* 64, 8–36.
- Harding, A.E. & Forrest, M.D. (1989) Analysis of multiple geological data sets from the English Lake District. *IEEE Transactions, Geoscience Remote Sensing* GE-27, 732–739.
- Harris, J.R., Renee, A.N., Ballantyne, B. & Sheridan, C. (1998) Mapping altered rocks using Landsat TM and lithochemical data: Sulphurets–Brucejack Lake district, British Columbia,

- Canada. *Photogrammetric Engineering and Remote Sensing* 64, 309–322.
- Horler, D.N.H., Barber, J. & Barringer, A.R. (1980) Effects of heavy metals on the absorbance and reflectance spectra of plants. *International Journal of Remote Sensing* 1, 121–136.
- Huntington, J.F. & Green, A.A. (1988) Recent advances and practical considerations in remote sensing applied to gold exploration. In: *Bicentennial Gold 88*, Melbourne.
- Irvine, J.M., Evers, T.K., Smyre, J.L., Huff, D., King, A.L., Stahl, G. & Odenweller, J. (1997) The detection and mapping of buried waste. *International Journal of Remote Sensing* 18, 1583–1596.
- Jansma, P.E. & Lang, H.R. (1996) Application of spectral stratigraphy to Upper Cretaceous and Tertiary rocks in southern Mexico: Tertiary graben controls on volcanism. *Photogrammetric Engineering and Remote Sensing* 62, 1371–1378.
- Jutz, S.L. & Chorowicz, J. (1993) Geological mapping and detection of oblique extensional structures in the Kenyan Rift Valley with SPOT/Landsat TM data merge. *International Journal of Remote Sensing* 14, 1677–1688.
- Keighley, J.R., Lynn, W.W. & Nelson, K.R. (1980) Use of Landsat images in tin exploration, Brazil. *Proceedings of the 14th Symposium on Remote Sensing of Environment, San Jose*, pp. 341–343.
- Kenea, N.H. (1997) Improved geological mapping using Landsat TM data, Southern Red Sea Hills, Sudan: PC and HIS decorrelation stretching. *International Journal of Remote Sensing* 18, 1233–1244.
- Kezheng, B. (1980) The application of remote sensing techniques to prospecting for metal deposits. *Proceedings of the 14th Symposium on Remote Sensing of Environment, San Jose*, pp. 135–142.
- Kierein-Young, K.S. (1997) The integration of optical and radar data to characterize mineralogy and morphology of surfaces in Death Valley, California, USA. *International Journal of Remote Sensing* 18, 1517–1542.
- Koch, M. & Mather, P.M. (1997) Lineament mapping for groundwater resource assessment: a comparison of digital Synthetic Aperture Radar (SAR) imagery and stereoscopic Large Format Camera (LFC) photographs in the Red Sea Hills, Sudan. *International Journal of Remote Sensing* 18, 1465–1482.
- Krishnamurthy, J. (1998) The evaluation of digitally enhanced Indian Remote Sensing Satellite (IRS) data for lithological and structural mapping. *International Journal of Remote Sensing* 18, 3409–3437.
- Kruse, F.A. (1988) Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in Northern Grapevine Mountains, Nevada and California. *Remote Sensing of Environment* 24, 31–52.
- Kruse, F.A., Kierein-Young, K.S. & Boardman, J.W. (1990) Mineral mapping at Cuprite, Nevada with a 63-channel imaging spectrometer. *Photogrammetric Engineering and Remote Sensing* 56, 83–92.
- Labovitz, M.L., Masuoka, E.J., Bell, R., Siegrist, A.W. & Nelson, R.F. (1983) The application of remote sensing to geobotanical exploration for metal sulphides—results from the 1980 field season at Mineral, Virginia. *Economic Geology* 78, 750–760.
- Labovitz, M.L., Masuoka, E.J., Bell, R., Nelson, R.F., Latsen, E.A., Hooker, L.K. & Troensegaard, K.W. (1985) Experimental evidence for spring and autumn windows for the detection of geobotanical anomalies through the remote sensing of overlying vegetation. *International Journal of Remote Sensing* 6, 195–216.
- Lattman, L.H. & Parizek, R.R. (1964) Relationship between fracture traces and the occurrence of groundwater in carbonate rocks. *Journal of Hydrology* 2, 73–91.
- Legg, C.A. (1990) *Remote Sensing, an Operational Technology for the Mining and Petroleum Industries*. Institute of Mining and Metallurgy, London.
- Liggett, M.A. & Childs, J.F. (1976) An application of satellite imagery to mineral exploration. *Proceedings of the 1st Annual William T. Pecora Memorial Symposium, Sioux Falls, South Dakota*, pp. 253–270.
- Loughlin, W.P. (1990) Geological exploration in the western United States by use of airborne scanner imagery. In: *Remote Sensing, an Operational Technology for the Mining and Petroleum Industries* (ed. C.A. Legg), pp. 223–241. Institute of Mining and Metallurgy, London.
- MacDonald, H.C., Clark, B.C., Taylor, D.S., Rainwater, G.L., Kaupp, V.H. & Waite, W.P. (1989) *Assessment of Shuttle Imaging Radar and Landsat Imagery for Ground-water Exploration in Arid Environments*. Miscellaneous Paper EL-79-6, U.S. Army Corps of Engineers, Washington DC.
- Marsh, S.E. & McKeon, J.B. (1983) Integrated analysis of high-resolution field and airborne spectroradiometer data for alteration mapping. *Economic Geology* 78, 618–632.
- Massonnet, D., Briole, P. & Arnaud, A. (1995) Deflation of Mount Etna monitored by spaceborne radar interferometry. *Nature* 375, 567–570.
- Memmi, J.M. & Pride, D.E. (1997) The application of Diamond Exploration Geoscientific Information System (DEGIS) for integrated diamond exploration in the north-central United States of America. *International Journal of Remote Sensing* 18, 1439–1464.
- Milton, N.M., Collins, W., Chang, S.-H. & Schmidt, R.G. (1983) Remote detection of metal anomalies on Pilot Mountain, Randolph County, North Carolina. *Economic Geology* 78, 605–617.
- Missalati, A., Prelat, A.E. & Lyon, R.J.P. (1979) Simultaneous use of geological, geophysical and Landsat digital data in uranium exploration. *Remote Sensing of Environment* 8, 189–210.
- Murphy, R.J. (1995) Mapping of jasperoid in the Cedar Mountains, Utah, USA, using image spectrometer data. *International Journal of Remote Sensing* 16, 1021–1041.
- Newton, A.R. & Boyle, T.P. (1993) Discriminating rock and surface types with multispectral satellite data in the Richtersveld, NW cape Province, South Africa. *International Journal of Remote Sensing* 14, 943–959.
- Norman, J.W. (1980) Causes of some old crustal failure zones interpreted from Landsat images and their significance in regional mineral exploration. *Transactions of the Institute Mining and Metallurgy (Section B: Applied Earth Science)* 89, B63–B72.
- Offield, T.W. (1976) Remote sensing in uranium exploration. In: *Exploration of Uranium Ore Deposits*, pp. 731–744. International Atomic Energy Agency, Vienna.
- Offield, T.W., Abbott, E.A., Gillespie, A.R. & Loguercio, S.O. (1977) Structural mapping on enhanced Landsat images of southern Brazil: tectonic control of mineralization and speculations on metallogeny. *Geophysics* 42, 482–500.
- Oppenheimer, C. (1998) Volcanological applications of meteorological satellites. *International Journal of Remote Sensing* 19, 2829–2864.
- Oppenheimer, C., Rothery, D.A., Pieri, D.C., Abrams, M.J. & Carrere, V. (1993) Analysis of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data of volcanic hot spots. *International Journal of Remote Sensing* 14, 2919–2934.

- Paradella, W.R., Bigneli, P.A., Veneziani, P., Pietsch, R.W. & Toutin, T. (1997) Airborne and spaceborne Synthetic Aperture Radar (SAR) integration with Landsat TM and gamma-ray spectrometry for geological mapping in a tropical rain forest environment, the Carajás Mineral Province, Brazil. *International Journal of Remote Sensing* **18**, 1483–1502.
- Peltzer, G. & Rosen, P. (1995) Surface displacement of the 17 May 1993 Eureka Valley, California, earthquake observed by SAR interferometry. *Science* **268**, 1333–1336.
- Peters, D.C. (1983) Use of airborne multispectral scanner data to map alteration related to roll-front uranium migration. *Economic Geology* **78**, 641–653.
- Plant, J.A., Jones, D.G., Brown, G.C., Colman, T.B., Cornwell, J.D., Smith, K., Walker, A.S.D. & Webb, P.C. (1988) Metallogenic models and exploration criteria for buried carbonate-hosted ore deposits: results of a multidisciplinary study in eastern England. In: *Mineral Deposits within the European Community* (eds J. Boissonnas & P. Omenetto), pp. 321–377. Springer-Verlag, Berlin.
- Podwysocki, M.H., Segal, D.B. & Abrams, M.J.* (1983) Use of multispectral scanner images for assessment of hydrothermal alteration in the Marysvale, Utah, mining area. *Economic Geology* **78**, 675–687.
- Prost, G. (1980) Alteration mapping with airborne multispectral scanners. *Economic Geology* **75**, 894–906.
- Raines, G.L. & Wynne J.C. (1982) Mapping of ultramafic rocks in a heavily vegetated terrain using Landsat data. *Economic Geology* **77**, 1755–1769.
- Ramasamy, S.M. & Balaji, S. (1995) Remote sensing and Pleistocene tectonics of Southern India peninsula. *International Journal of Remote Sensing* **16**, 2375–2391.
- Ramos, V.A. (1977) Basement tectonics from Landsat imagery in mining exploration. *Geologie en Mijnbouw* **56**, 243–252.
- Resmini, R.G., Kappus, M.E., Aldrich, W.S., Harsanyi, J.C. & Anderson, M. (1997) Mineral mapping with Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor data at Cuprite, Nevada, USA. *International Journal of Remote Sensing* **18**, 1553–1570.
- Richards, D.M., Jones, V.T., Matthews, M.D., Maciolek, J., Pirkle, R.J. & Sidle, W.C. (1986) The 1983 Landsat soil-gas geochemical survey of Patrick Draw area, Sweetwater County, Wyoming. *Bulletin, American Association of Petroleum Geologists* **70**, 869–887.
- Rivard, B. & Arvidson, R.E. (1992) Utility of imaging spectrometry for lithologic mapping in Greenland. *Photogrammetric Engineering and Remote Sensing* **58**, 945–949.
- Rothery, D.A., Francis, P.W. & Wood, C.A. (1988) Volcano monitoring using short-wavelength infrared data from satellites. *Journal of Geophysics Research* **93**, 7993–8008.
- Rowan, L.C. & Bowers, T.L. (1996) Remote mineralogic and lithologic mapping of the Ice River alkaline complex, British Columbia, Canada, using AVIRIS data. *Photogrammetric Engineering and Remote Sensing* **62**, 1379–1385.
- Rowan, L.C., Wetlaufer, P.H., Goetz, A.F.H., Billingsley, F.C. & Stewart, J.H. (1974) Discrimination of rock types and detection of hydrothermally altered areas in South-central Nevada by the use of computer-enhanced ERTS images. *U.S. Geological Survey Professional Paper* **883**.
- Sabins, F.F. (1999) Remote sensing for mineral exploration. *Ore Geology Reviews* **14**, 157–183.
- Saraf, A.K., Cracknell, A.P. & MacManus, J. (1989) Geobotanical applications of airborne Thematic Mapper data in Sutherland, NW Scotland. *International Journal of Remote Sensing* **10**, 545–555.
- Schmidt, R.G. (1976) Exploration for porphyry copper deposits in Pakistan using digital processing of Landsat-1 data. *U. S. Geological Survey Journal of Research* **4**, 27–34.
- Schmugge, J.T. (1983) Remote sensing of soil; moisture: recent advances. *IEEE Transactions, Geoscience Remote Sensing* **GE-21**, 336–344.
- Segal, D.B. (1983) Use of Landsat multispectral scanner data for the definition of limonitic exposures in heavily vegetated areas. *Economic Geology* **78**, 711–722.
- Short, N.M. & Blair, R.W., Jr (1986) *Geomorphology from Space*. NASA SP-486, U.S. Government Printing Office, Washington, DC.
- Siegal, B.S. & Goetz, A.F.H. (1977) Effect of vegetation on rock and soil type discrimination. *Photogrammetric Engineering and Remote Sensing* **43**, 191–196.
- Simpson, C.J. (1978) Landsat: developing techniques and applications in mineral and petroleum exploration. *BMR Journal of Australian Geological Geophysics* **3**, 181–191.
- Spatz, D.M. (1997) Remote sensing characteristics of the sediment- and volcanic-hosted precious metal systems: image selection for explorations and development. *International Journal of Remote Sensing* **18**, 1413–1438.
- Vincent, R.K. (1977) Uranium exploration with computer processed Landsat data. *Geophysics* **42**, 536–541.
- Vincent, R.K. (1994) Remote sensing for solid waste landfills and hazardous waste sites. *Photogrammetric Engineering and Remote Sensing* **60**, 979–982.
- Vogel, A. (1985) The use of space technology in earthquake hazard assessment. *Proceedings of the 18th International Symposium on Remote Sensing Environment, Paris*.
- Wang, Q.M., Nishidai, T. & Coward, M.P. (1992) The Tarim Basin, NW China: formation and aspects of petroleum geology. *Journal of Petroleum Geology* **15**, 5–34.
- Whitney, G., Abrams, M.J. & Goetz, A.F.H. (1983) Mineral discrimination using a portable ratio-determining radiometer. *Economic Geology* **78**, 688–698.
- Williams, R.S. (ed.) (1983) Geological applications. In: *Manual of Remote Sensing* (ed. R.N. Colwell), 2nd edn, pp. 1667–1953. American Society of Photogrammetry, Falls Church, Virginia.
- Windeler, D.S. (1993) Garnet-pyroxene alteration mapping in the Ludwig skarn (Yerington, Nevada) with Geoscan airborne multispectral data. *Photogrammetric Engineering and Remote Sensing* **59**, 1277–1286.

Appendix A | Stereometry

Viewing remotely sensed images in three dimensions relies on the stereoptic capabilities of the human visual system. As outlined in Section 2.4, for an object to be perceived in three-dimensional space the absolute parallax of its images on the retinas of both eyes must be greater than the minimum resolution of the visual system. For two objects to be discriminated in three dimensions similarly means that their relative parallax or parallax difference at the eyes also must exceed this minimum. If the positions of an imaging platform at high altitude or in orbit are spaced widely enough during the capture of two overlapping images, then the resulting absolute and relative parallaxes on the images can be large enough for them to be resolved visually. This being the case, a stereoptic model of the surface topography is seen if each image can be viewed separately by each eye. The adjacent images are substituted for the natural views using both eyes, and the effective eye base is increased from 6 or 7 cm to as much as 100 km. Relief is perceived because a continuum of parallax difference exists for all of the points that constitute the surface.

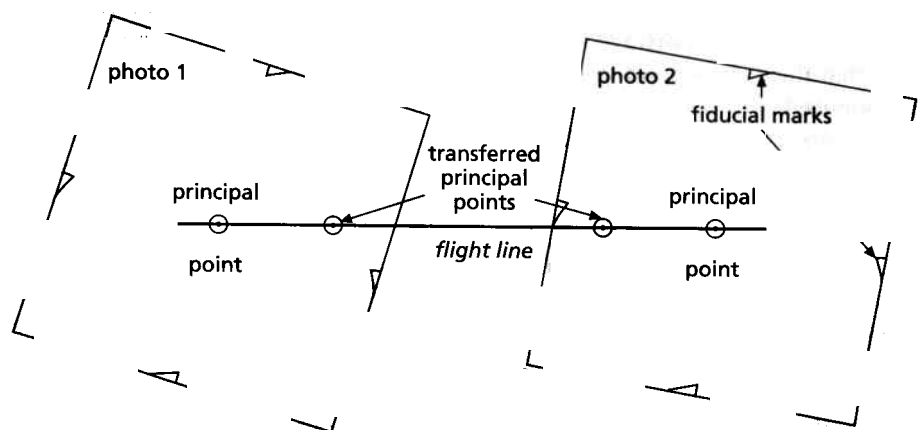
Parallax is normally achieved in images by different positions of the platform as in overlapping aerial photographs taken from adjacent positions along the same flight line, or from different flight lines with different viewing geometries, as in the case of radar and line-scan or pushbroom systems (Chapters 3 and 7). Landsat images from adjacent paths have between 15 and 80% overlap, from the Equator to the highest latitudes covered by the system. Stereoptic potential is available from both the SPOT and MOMS systems (Section 3.12) by selective pointing off-nadir and a combination of forward- and nadir-looking devices, respectively. It also can be introduced synthetically into a single image

from other two-dimensional data, such as those for digital elevation or geophysical and geochemical variables (Chapter 8). When the unchanged image is viewed with one eye and the transformed image with the other, the result is a synthetic stereoptic model of the artificial topology. For both kinds of image pair the trick is to persuade the visual system that the eyes are being used in a normal way.

For two overlapping aerial photographs to be viewed stereoptically, they must first be orientated so that the eye base is parallel to the flight direction. Secondly, the traces of the flight time on both images must coincide. This is necessary because of radial relief displacement and distribution of parallax parallel to the flight direction. This means identifying the principal points of both images on each image and aligning the four points (Fig. A1). For line-scan and pushbroom images that overlap, and radar images that contain parallax differences, apparent relief displacement and relative parallax between image swaths are parallel to the scan lines and the look direction, respectively. There is neither relative parallax nor relief displacement parallel to the flight direction. The images must be arranged so that the two flight lines are parallel. Finally, the separation of the two images must be adjusted to suit the method of viewing.

The normal reaction to unaided viewing of two separate objects at close range is for the eyes to converge on only one of the objects at one time. For the untrained observer therefore only one of a pair of images can be seen at one time. Stereoptic fusion of the two images means overcoming the natural tendency for close-range convergence. When looking into the far distance the eyes assume a parallel orientation of their optic axes, as well as the necessary focal length for clear vision. One

Fig. A1 This figure shows how an overlapping pair of aerial photographs is set up for stereoptic viewing using a reflecting stereoscope. The principal point of each photograph is located by the crossing of lines ruled between the fiducial points on the margin of the frame. The objects on the ground to which the principal points are related are then located on the adjacent photographs to give the four points that define the flight line and its direction. Separation of the images is determined by the amount of overlap and the geometry of the stereoscope.



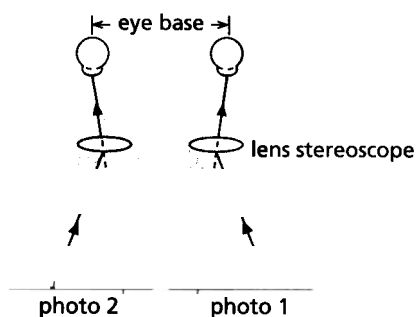


Fig. A2 This simple ray diagram shows how the lenses of a stereoscope optically straighten the line of sight from each eye when the eyes are focused at and converge towards a point at the normal distance for comfortable reading. The effect is to enable each eye to see a separate image and to create the illusion of viewing the stereoptically fused pair of images from a great distance.

method of achieving unaided stereopsis of an image pair is to stare 'through' them, as in a daydream, and then to coax the eyes to regain focus without convergence. With luck, a blurred stereoptic model is achieved, which becomes sharper with perseverance. Of course, to accomplish this means that common parts of both images must have roughly the same separation as the eyes. A simple aid for the unsuccessful beginner is to separate the two views physically with a card held at the bridge of the nose. Once fusion and focus can be achieved, then the card can be dispensed with. As useful as unaided stereoptic viewing of image pairs is, only a minority of people can manage to do it. The majority require some kind of optical assistance—a stereoscope.

In a simple lens stereoscope (Fig. A2) the lenses refract rays from an object portrayed on both images in such a way that they correspond to the convergence normally experienced with close-range vision. The two images of the object are focused on the left and right retinas with no strain, and the stereoptic model of relief quickly springs into being, as can be tested with any of the stereopairs of images in Chapters 4, 7 and 8. The problem is that only a portion of the overlap between the two images as wide as the eye base can be viewed. To view the full area of overlap means using a system of lenses together with prisms or mirrors, or both, in a reflecting stereoscope (Fig. A3). This combines the refraction of rays to the normal convergence angle with an increase in the effective eye base, in much the same way as

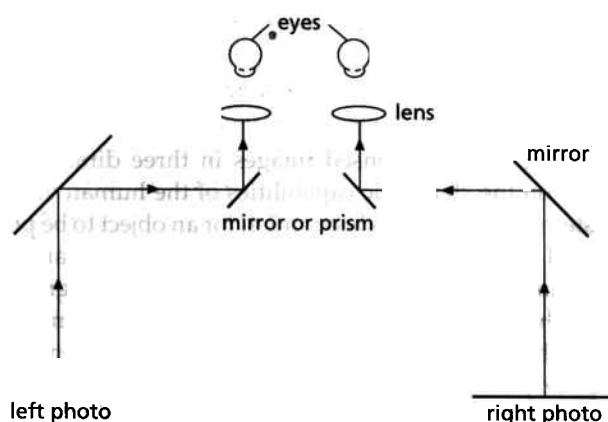


Fig. A3 The lens system in a reflecting stereoscope optically straightens the line of sight, as in a lens stereoscope. The prisms and mirrors enable the pair of images to be separated sufficiently for the whole area of overlap to be viewed stereoptically.

binoculars. All stereoscopes, including those incorporated in sophisticated plotting and measuring instruments, are based on these simple principles.

The stereoptic potential of image pairs with parallax differences also can be exploited by other optical principles. One involves projecting both images on to a screen, one through a filter that polarizes light in the horizontal plane, the other through a vertically polarizing filter. The screen is viewed through spectacles with filters having corresponding polarities. Each eye receives light from only one of the two projected images, and stereoptic fusion results. This is a handy method of presenting and teaching the principles of stereoptic photo-geological interpretation. Another way of ensuring that each eye sees only one of a pair of images is to print or project the images in separate primary colours. Viewing with spectacles having filters with the same arrangement of primaries ensures that only light from the left image is received by the left eye, and vice versa. Coloured light from the other image is absorbed by the filter. This is the principle of the anaglyph, once popular in children's books and three-dimensional movies. All the stereoscopic figures in Chapter 4 are presented as anaglyphs on the CD, which has an accompanying pair of red-green spectacles.

Viewing most of the pairs of images in Chapter 4 with a lens stereoscope reveals a marked degree of vertical exaggeration of the topographies involved. This phenomenon is very useful for highlighting subtle topographic features that otherwise would go unnoticed. However, it also causes all slopes to appear steeper than they really are. This is particularly irritating when dips of geologically important surfaces need to be estimated. In most cases vertical exaggeration means that estimates can be made only in terms of horizontal, gentle, moderate or steep dip. The amount of vertical exaggeration

depends on the geometries of both the imaging and viewing systems.

To ensure complete stereoptic cover aerial photographs generally have a 60% overlap or more along a flight line. Two factors contribute to the overlap. First, the focal length of the lens governs the field of view and, for a fixed frame size, the scale of the image. The shorter the focal length, the wider the field of view and the smaller the scale. The distance flown by the platform between successive exposures—the air base—is the obvious second factor. At a fixed altitude therefore 60% overlap is achieved by a larger air base for a lens with short focal length than for one with a longer focal length. The two images in the first case contain a greater relative parallax for two objects with different elevations. This produces an enhanced three-dimensional effect when the two images are viewed stereoptically. From these simple relationships the degree of vertical exaggeration can be deduced to be directly proportional to the ratio between the air base and the altitude.

Because of the advantages of vertical exaggeration for interpretation, and because slope angles can be measured by other methods, aerial photographs usually have an air base : altitude ratio of between 0.4 and 0.6. This gives exaggeration factors of between three and four times normal. Beyond exaggeration factors of 4, the amount of relative parallax between objects with high and low elevation may be larger than the visual system can accommodate. To view the whole scene stereoptically requires different image separations for the low and high parts. The field of view of orbiting systems, although large in absolute terms, is small relative to altitude to ensure that distortions at the margin of images are kept to a minimum. As the equivalent of air base for overlapping Landsat images decreases towards the poles from a maximum at the Equator, the vertical exaggeration changes from about 1.2–0.2. A useful stereoptic model is only possible for latitudes below about 50°. The stereoscopic imaging arrangements for SPOT and MOMS aim at exaggeration factors of around two.

The parallax differences between overlapping images provide the basis for quantitative measurements of the relative elevations of objects. Where b is the air base of the images, H is the altitude of the platform and Δp is the parallax difference between the images of two objects, their vertical separation h is given by:

$$H = H\Delta p/b \quad (\text{A1})$$

provided measurements of Δp and b are in the same units. Converting the parallax differences between objects on pairs of overlapping radar images to relative elevations involves different equations, which depend on the particular configuration of the sensors. For the commonly used same-side, same-height arrangement the elevation difference h is given by:



Fig. A4 Viewing these converging rows of dots with a lens stereoscope produces a stereoptic illusion that they form a single row which dips into the page. This is the principle behind both the parallax wedge and the stereometer. In a parallax wedge, each pair of dots is accompanied by a scale division indicating the absolute parallax between the dots. In a stereometer this distance is measured using a micrometer to which the plates carrying the dots are attached.

$$h = \Delta p / (\tan \theta_2 - \tan \theta_1) \quad (\text{A2})$$

where θ_1 and θ_2 are the depression angles from the antenna to objects on the two images.

Because parallax differences in most images are small, they cannot be measured precisely using a ruler. It is more convenient and useful to use the principles of stereopsis during stereoptic viewing itself. The simplest method is to use a transparent plate engraved accurately with two converging rows of dots, crosses or other marks, known as a parallax wedge (Fig. A4). When viewed with a lens stereoscope the two converging rows of marks appear as a line dipping through the stereoptic model, because of the regular variation in absolute parallax between adjacent marks. By moving the wedge over the stereopair at right angles to the flight line, one of the fused pairs of marks can be 'sat' on a particular topographic feature. The corresponding absolute parallax of the two marks is read from the scale. Another pair of marks is fused so the floating mark 'sits' on a feature at a different elevation, to give the absolute parallax at that point. The difference between the two readings is then used in Equation (A1) to give a difference in elevation between the two points. Clearly, a parallax wedge together with a lens stereoscope is a useful tool in the field.

The parallax bar or stereometer uses the same principles as the parallax wedge. It consists of two small transparent plates, each engraved with a dot or a cross, which are linked by a micrometer screw so that their separation can be measured precisely. With the bar aligned parallel to the flight direction the separation of the spots is adjusted so that the fused spot 'sits' on the surface, first at one position and then at another. The parallax difference is read from the micrometer, and is converted to a difference in elevation.

More sophisticated instruments used in stereometry incorporate means of calculating elevation differences, together with mechanisms for removing tilt produced in the stereopairs during flying and for transferring information from the images to a base map. A pair of marks similar to those used in the parallax bar are on linked arms, so the fused mark can be moved along topographic or geological features. Another linkage incorporating the appropriate scale conversion drives a pencil across a base map to record the features. If the fused mark is maintained apparently on the surface, then it is possible to map topographic contours directly. Further refinement combines the results from such plotters and measuring systems through microprocessors to produce digital maps. Further details of the various instruments used in photogrammetry are beyond the scope of this book.

The angle between the horizontal and a line that joins two points can be defined by the ratio between their difference in elevation and their horizontal separation. This ratio is the tangent. The angle therefore can be estimated by stereometric measurement of parallax difference for the two points and their horizontal separation. If the two points lie on the same dipping surface, such as a topographic slope or a bedding plane, then this angle is the true dip of the surface. The advantages of this possibility for the geologist equipped with aerial photographs, a stereoscope and a parallax wedge or bar are obvious. One in particular is that the minor variations in dip that plague field measurements are removed to give a local average.

Relative elevations of objects on images can be measured without involving parallax, but with less precision by using the lengths of shadows that they cast. If the angle θ at which the object is illuminated by the Sun or by radar is known, then the elevation of the top of the object above its base h is related to the length of its shadow l by:

$$h = l \tan \theta \quad (\text{A3})$$

Useful as this method may be for estimating topographic elevations of landforms, it is of little use in estimating dips, because the very presence of shadows obscures geological features.

Automated derivation of elevation data from digital stereopairs of images uses the basic principles of stereometry. One of the images is geocoded using the geographical co-ordinates and elevations of ground control points. The operator creates common tie points on the

two images to place them both in the same frame of reference. This involves rotation and aligning the images along the flight path that connects the two principal points, as well as resampling them to the same pixel dimensions. This is much the same as the set-up for viewing with an optical stereoscope (Fig. A1). In a similar fashion to the eyes' adaptation to parallax in creating stereoptic fusion, the software finds the parallax in the along-flight direction to calculate elevation differences for every common pixel. Calibrated to elevation values for control points, this gives a DEM with absolute elevation above the sea-level datum. The difference lies in the software using pattern-matching algorithms to 'recognize' common points automatically. The efficiency with which it does this depends on the images' quality. This means that a high percentage of pixels are not recognized and omitted from the calculations. Instead, those points for which parallax can be measured form the basis of a TIN (Section 8.2.4), which the software interpolates to a continuous topographic surface. Consequently, we can see details in stereopairs that do not appear in DEMs.

Further reading

- Coupland, D.H., Vincent, R.K. & Pleitner, P.K. (1985) Examples from a new digital elevation mapping algorithm for stereoscopic image data. *Proceedings of the 18th International Symposium Remote Sensing of Environment, Paris*.
- Day, T. & Muller, J.-P. (1989) Digital elevation model production by stereomatching SPOT image pairs: a comparison of algorithms. *Image and Vision Computing* 7, 95–102.
- Lattman, L.H. & Ray, R.G. (1960) *Aerial Photographs in Field Geology*. Holt, Rinehart and Winston, New York.
- Orti, F. (1981) Optimal distribution of control points to minimize Landsat image registration errors. *Photogrammetric Engineering And Remote Sensing* 47, 101–110.
- Rochon, G., Haja, S.R. & Simard, R. (1985) Automatic digital elevation model extraction from digital stereo images. *Proceedings of the 18th International Symposium Remote Sensing of Environment, Paris*.
- Shlien, S. (1979) Geometric correction, registration and resampling of Landsat imagery. *Canadian Journal of Remote Sensing* 5, 74–89.
- Slama, C.C. (ed.) (1981) *Manual of Photogrammetry*, 4th edn. American Society of Photogrammetry, Falls Church, Virginia.
- Thompson, M.H. (ed.) (1966) *Manual of Photogrammetry*, 3rd edn. American Society of Photogrammetry, Falls Church, Virginia. (2 Volumes.)
- Wolf, P.R. (1974) *Elements of Photogrammetry*. McGraw-Hill, New York.

Appendix B Image Correction

Most kinds of remote-sensing device produce unavoidable geometric distortions in the images that they produce. These departures from normal map projections of the Earth's surface include the warping effects of deviations in the flight direction and attitude of the platforms for remote sensing (Chapter 3). As well as distortions, defects in the radiometric fidelity of the images are also common. These range from mistakes in the exposure, processing and printing of photographic film to the results of systematic foibles in digital imaging devices.

Section 5.2 covered poor contrast resulting from atmospheric scattering and compression in digital images, and how to compensate for them. Other common blemishes in digital images are caused by malfunctions in the hardware, or are peculiarities of the method itself. The manufacture of line-scanning devices and, to a lesser extent, pushbroom systems fails to achieve a perfect match between the responses of individual detectors to the same stimulus. Each detector may also drift or deteriorate with time, especially if its temperature fluctuates. The net result is striping parallel to the scan lines on the image. In the Landsat MSS and TM instruments detectors for visible and near-infrared radiation are mounted in banks of 6 and 16, respectively. The effect of mismatches between the detectors is repeated patterns of striping, each stripe being 6 or 16 lines deep.

Brief failures in the circuits of a detector result in spurious signals during a scan cycle. This may produce lines without data or, where the data have uniformly high values, which appear as black or white line drop-outs, respectively. Build-up of minute electrostatic charges in the circuits of the system and their eventual discharge can result in a spurious response for a single pixel. This noise effect generally is distributed as random speckles in an image. A very similar result stems from the complex returns of microwave signals, so that radar images are systematically infested with speckle. In fact, the speckle does contain valid information relating to attributes of the surface, but this is rarely of much interest to the geologist, and it disturbs the topographic patterns of interest.

Not only do defects degrade the visual quality of an image, but many of them have spatial frequencies close to where the achromatic MTF of the human visual system rises to a maximum (Section 2.2). We notice them more readily than natural variations, and even a slight defect disturbs interpretation.

B.1 Geometric rectification

All images contain geometric distortions of some kind. There are many contributory effects—different kinds of platform instability, optical and mechanical distortions in the radiation-gathering set-up, Earth rotation beneath the platform, and both parallax and scale effects resulting from topographic relief. Apart from the last two, simple corrections of the geometry remove most distortions. Where precise information about platform attitude and track relative to the Earth's rotation is available, this can be built into automatic correction programmes. In polar-orbiting satellite images the correction for Earth rotation is simply a systematic line-by-line displacement of pixels, very much like giving a slant to the edge of a pack of cards.

At large scales the Earth's surface approximates a plane on which the topography is superimposed. Rectification of an image essentially restores objects to their correct relative disposition in two-dimensional space as if they were all viewed from directly above. Small-scale images from metsats contain distortion effects resulting from Earth's curvature. They are projections of a spherical surface into two dimensions. All maps have similar distortions. Different scales and different uses call for different kinds of projection, however, such as stereographic, conical, cylindrical and more complex renditions. All aim to maintain a common scale over a whole map, a minimum of distortion, and a uniform representation of shapes and areas. As maps are the stock-in-trade for all users of remotely sensed images, geometric manipulations of images attempt to register images to the map projection required for a particular project.

Map projections involve two main features: a geometrical model that closely approximates the shape of the Earth, and a co-ordinate system for referring to locations. Bulging because of rotation means that the Earth is slightly flattened—its polar radius is slightly shorter than that to points on the Equator. The planet is close to being an ideal ellipsoid. However, topography and variations in density within the Earth result in local changes in its gravitational field. For surveyors this is something of a problem, because defining a horizontal surface depends on the angle of a plumb line! The imaginary surface at which gravity is the same as at mean sea-level—the geoid—is not an ellipsoid, and can deviate from the ideal by up to 100 m.

Co-ordinates that express map position always refer to a geodetic datum on which the map is based. Such a

datum uses an ellipsoidal surface, but no single one fits every part of the planet. So different areas need different ellipsoids, and a datum specifies both the ellipsoid appropriate for a region of Earth's surface and the origin for its co-ordinates. In some co-ordinate systems, such as the metric grid used in the Universal Transverse Mercator system, different ranges of latitude and longitude constitute different zones. Finally there are projections. The object of a projection is to render the ellipsoidal Earth's surface as another surface, which can be flattened without tearing or distortion on to a flat piece of paper. Only three kinds of surface can do this: planes that lie tangential to the surface, cones and cylinders. These form the basis for the many projections that cartographers use. Each involves distorting surface features in some way; in terms of scale, area or shape. The first is impossible to achieve everywhere, and the second and third, although possible, are mutually exclusive. Such is the life of the geographer! However, maps are what we use to refer to land and its geology; they are the arbiters of accuracy. That is why transforming images to conform with maps is called rectification! Things go awry when co-ordinates refer to inappropriate datums, and that can lead to errors of up to almost a kilometre.

As most digital images consist of rectilinear arrays of DN for columns and rows of pixels, rectification consists of a transformation from one set of Cartesian co-ordinates to another. The three-dimensional co-ordinates of points, expressed as latitude and longitude, can be expressed by reference to a rectilinear array on maps. However, describing a position accurately in degrees, minutes and seconds of arc is difficult. Points in a metric grid of eastings and northings are more easily assessed. The most commonly used grid of this kind is the Universal Transverse Mercator (UTM) projection found on most modern maps.

Image distortions, except those caused by parallax due to differences in surface elevation, can be removed by registration of a digital image to a map base. The process begins by identifying topographic or cultural features visible on both the image and the map. The map co-ordinates are assigned to the pixel and line co-ordinates of a number of these ground control points on the image. A very similar procedure can be used to register an image of one type to another. Depending on the number of control points, a pair of polynomial equations, such as:

$$X_m = a + bX + cY + dX^2 + eXY + fY^2 + gX^2Y + hXY^2 + iX^3 + jY^3 + \dots$$

$$Y_m = z + yX + xY + wX^2 + vXY + uY^2 + tX^2Y + sXY^2 + rX^3 + qY^3 + \dots$$

where X_m and Y_m are map co-ordinates and X and Y are image co-ordinates, can be generated. These equations

describe the deviation in two dimensions of the image from the map projection. The pair of polynomial transforms allow a computer to warp the image to register with the map, by treating the image mathematically like a rubber sheet. In the production of a new rectified array of pixels for the image, it is possible to change the dimensions and shapes of pixels. This can involve shifting, rotating and distorting the original pixels.

Each new pixel in a rectified image must have assigned to it a DN derived from those in the original image. This is resampling. As the new pixels are displaced relative to the raw array of DN, simply assigning the original DN of a shifted pixel will produce errors in the new image. The simplest means of allowing for this effect is to assign the DN of the nearest raw pixel to a pixel in the new array. This is the nearest-neighbour method (Fig. B1a). Being a crude approximation it results in the resampled image appearing 'blocky'. It retains the original statistics of the data, however, and thus is commonly used when the resampled image is intended for multispectral classification. A more geometrically precise approach is to assign the distance-weighted mean DN of several raw pixels immediately surrounding the position of a rectified pixel using convolution (Section 5.3). Best results involve the mean of the 16 surrounding raw pixels using a method known as cubic convolution (Fig. B1b). This reduces the blockiness in the rectified image, and keeps disturbance of the original frequency distribution of DN to a minimum.

The parallax and scale variations that cannot be removed from an image by its digital registration to a map depend on both elevation differences and position

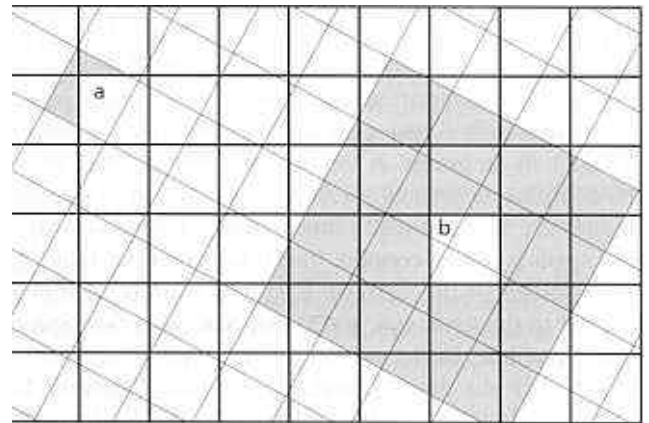


Fig. B1 In the nearest-neighbour method of resampling the digital number (DN) for an output pixel (a) is taken unchanged from the input pixel (shaded) lying nearest to the new one. Resampling by cubic convolution produces an output DN at b, which is the mean value of the DN from those 16 input pixels grouped around the position of the output pixel (shaded). A more sophisticated version weights the contribution of the DN of each input pixel to the mean according to its distance from the new pixel position.

of features relative to the nadir of the imaging system. For most satellite images these do not produce distortions that are much greater than the imprecision inherent in surveying and printing small-scale maps. For instance, an elevation difference of 1 km at the edge of a Landsat image gives a displacement equivalent to only 160 m. Parallax does introduce important distortions in aerial photographs with scales greater than 1 : 100 000.

To overcome parallax and scale distortions, interpretations from stereopairs of aerial photographs are transferred to maps using various optical-mechanical plotters. Where maps do not exist at a scale suitable for the interpretation they must be constructed from the photographs using triangulation methods that allow for radial-relief displacements. The parallax involved in radial-relief displacement can be removed from aerial photographs using an orthophotoscope. In this instrument the image is rephotographed through a narrow slit. The operator views the stereomodel and attempts to keep an image of the slit at the ground surface (see Appendix A) as the image is scanned in one direction as a series of narrow swaths. Through various mechanical linkages this manipulation is translated into up-and-down movements of the film on to which the image is projected. This controls the local scale of the new image on the film. The process can involve direct rephotography, or the three-dimensional co-ordinates controlling the film and slit positions can be stored and replayed through the orthophotoscope when the task is complete. The final product is an orthophotograph, which represents each point as though it has been viewed from directly above, just as in a map. Where digital elevation data are available complex computer programs can remove parallax from digitized photographs, again resulting in an orthographic image.

B.2 Replacing dropped lines

Figure B2(a) illustrates graphically the disturbing effects of a relatively few lines in images that contain spurious data. The problem can be reduced by replacing each line with data derived from those adjacent to it. Each pixel in the dropped line is assigned the DN of the mean value of the pixels immediately above and below it. Although the new line is entirely artificial, it is so similar to its neighbour that it becomes barely noticeable, except under high magnification. The degree of cosmetic improvement that can be achieved is clear from Fig. B2(b).

B.3 De-striping

Where image striping stems from different responses from each detector in a bank, as in Landsat MSS and

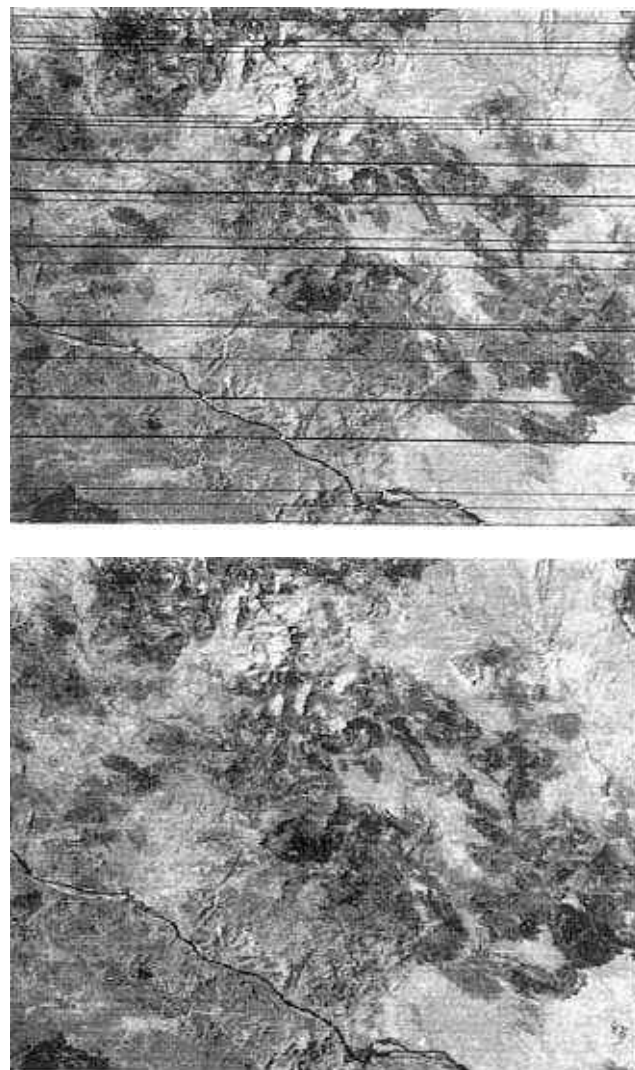


Fig. B2 The Landsat-2 MSS image of an area in the Ethiopian Highlands (a) is severely affected by lines of spuriously low data. In image (b) the spurious lines have been replaced by data derived from those in the adjacent lines. At the scale the defect is hardly noticeable.

TM images, it appears as a regularly repeated defect throughout an image. A simple and rapid means of cosmetic treatment of such simple striping begins by comparing histograms for all of the detectors in a bank over a representative part of the scene. The histogram with the highest mean DN and the greatest variance is selected as the standard, and the others are then normalized to it. Individual DN for pixels contributing to each histogram are changed to fit the 'master' histogram using look-up tables (Chapter 5).

Histogram-matching methods remove stripes effectively from images of relatively homogeneous scenes. Where scenes contain large areas with very different overall brightness, the correction may leave residual stripes

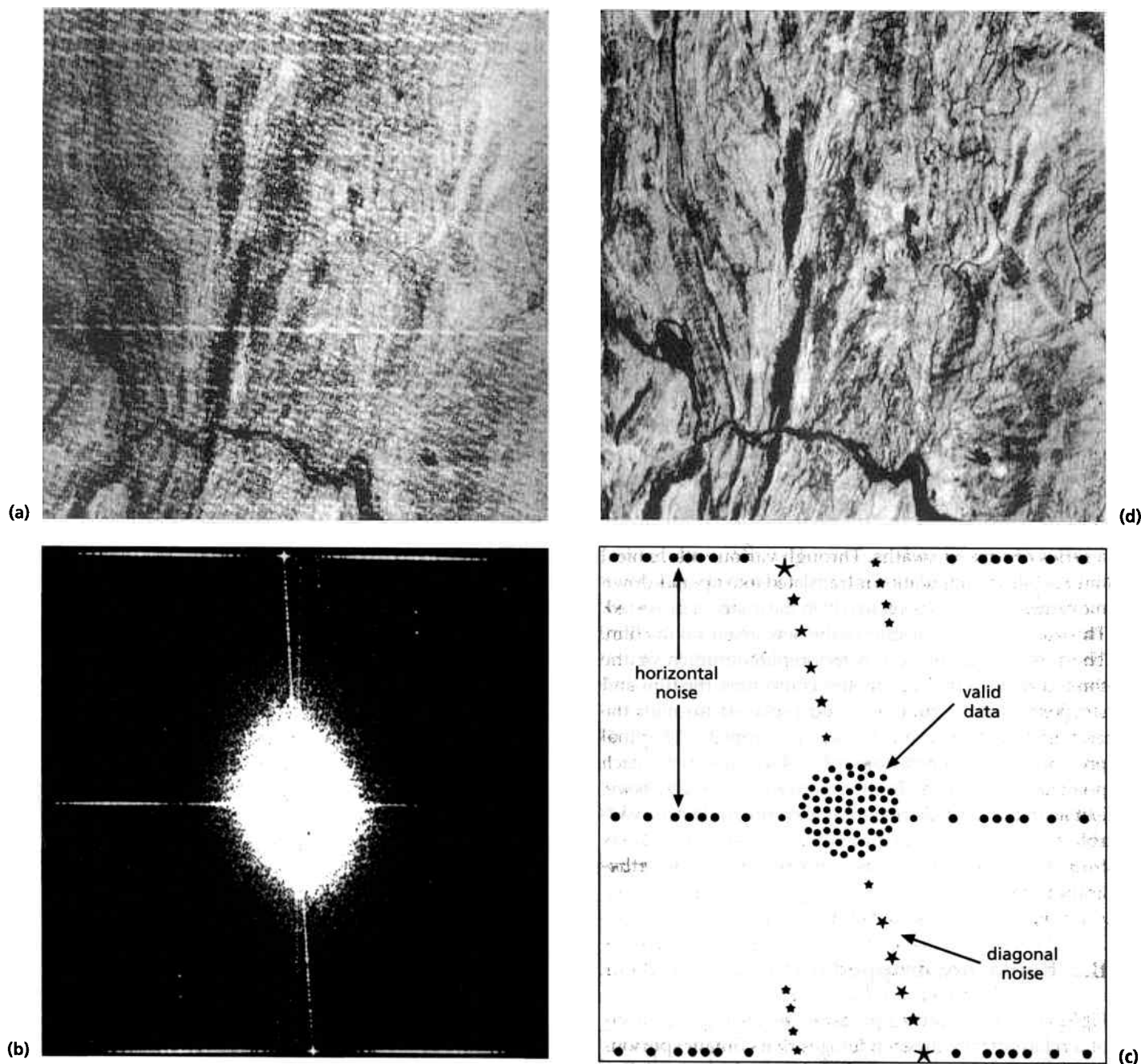


Fig. B3 A variety of problems in the JERS-1 short-wavelength infrared (SWIR) sensors produced severe defects in image (a) of band 8 covering part of the Eritrean Highlands. (b) The magnitude of the Fourier transform of image (a), in which the noise shows as a variety of 'spikes' and lines, as explained in

(c). Using filters specially designed for this unique set of defects almost completely removes the noise, as shown in the restored image (d). Courtesy of Beto de Souza Filho, University of Campinas, Brazil.

in particularly dark and light areas. More comprehensive de-striping methods involve more complex computations. One is to use convolution filtering based on matrices with dimensions that match the number of sensors in the imaging system. The filter effectively smooths out the striping. Other methods rely on identifying the

frequency and amplitude of striping and other all-pervasive, systematic defects, which are superimposed on the real variations, by analysing a Fourier transform of the image. The periodic noise can then be identified as a signal and suppressed by filtering in the frequency domain. The inverse transform to the spatial domain

then restores the cleaned-up image. Figure B3 shows the complex noise to which JERS-1 SWIR images are prone. Without cosmetic treatment they are difficult to interpret. In the frequency domain the noise is isolated as distinct signals, and almost wholly removed by specially designed filters that operate in the frequency domain.

B.4 Removal of random noise

Random noise, or speckle, degrades the visual impact of the raw images (see Fig. B5a), and therefore should be removed by some means. Being randomly distributed, it is uncorrelated with the real data in a multispectral image. In the decorrelation process involved in principal component analysis such poorly correlated data end up in the highest order components. Performing PCA, discarding the highest PC, and then reconstituting the data by the inverse transform is a logical, although by no means quick, solution to the visual disturbance introduced by noisy pixels.

For some noisy images, such as those produced by band ratioing techniques or by radar systems, principal component analysis is not an appropriate means of noise removal. The interesting information is best retained in the original data set. A hypothetical DN resulting from noise is represented in Fig. B4 by the number 255, and is surrounded by valid DN. Calculation of the mean of the nine DN to replace the DN of the central noisy pixel using a convolution filter produces a value of 51, and also blurs the image. The new pixel is still sufficiently different from its neighbours to appear as a speck in the image. However, the median DN of the sequence 24, 25, 26, 26, 27, 27, 255 is 26. As this is more representative of the surrounding DN, the resulting image is effectively cleared of noise, whereas real data remain almost unchanged. Such median filtering is possible using a 3×3 square convolution matrix in the spatial domain. Figure B5(b) shows the efficiency of median filtering.

Most image processing software packages include a range of filtering methods to remove image defects. Deciding on the most appropriate for a particular problem is a matter of trial and error.

| | | |
|----|-----|----|
| 24 | 26 | 25 |
| 27 | 255 | 26 |
| 26 | 25 | 27 |

Fig. B4 This array of digital numbers (DN) for a 3×3 part of an image consists of eight valid DN surrounding a spuriously high DN, which constitutes system noise. Replacing the value of 255 with the median DN for all nine pixels removes the noise effect.



Fig. B5 An image of the ratio between Landsat MSS bands 4 and 5 for an area in the Chilean Andes (a) is plagued by an increase in the signal-to-noise ratio, which is inevitable in ratio images. By convolving the raw image with a 3×3 median filter the speckle and residual striping is virtually removed (b).

Associated resources on the CD-ROM

You can access the resources by opening the HTML file Menu.htm in the Imint folder on the CD using your Web browser. More detailed information about installing the resources is in Appendix C.

Exercise 6 includes TNTlite processes that are useful in cosmetic treatment of image defects, and shows some of the artefacts that appear when using spatial filtering. Exercise 7 covers georeferencing and rectification of images.

Further reading

Bernstein, R. (eds) (1983) Image geometry and rectification. In: *Manual of Remote Sensing*, 2nd edn (ed. R.N. Colwell), pp. 873–922. American Society of Photogrammetry, Falls Church, Virginia.

Billingsley, F.C. (1983) Data processing and reprocessing. In: *Manual of Remote Sensing*, 2nd edn (ed. R.N. Colwell), pp. 719–792. American Society of Photogrammetry, Falls Church, Virginia.

Crippen, R.E. (1989) A simple spatial filtering routine for the cosmetic removal of scan-line noise from Landsat TM P-tape imagery. *Photogrammetric Engineering and Remote Sensing*. 55, 327–331.

Appendix C The CD-ROM Resources

The CD-ROM at the rear of the book contains resources aimed at extending and applying what you should have grasped from reading the text. I hope that it transforms *Image Interpretation in Geology* from a textbook supporting courses in geological remote sensing to a course in its own right, that will help professional geologists develop new practical skills through home study, and can support college instructors in enriching existing courses or setting up new ones for undergraduate geologists.

The central component uses a version of the professional Map and Image Processing System (TNTmips) software developed and marketed by MicroImages Inc. of Lincoln, Nebraska, USA. MicroImages are the only software developers in this field to offer a freeware version (TNTlite), which is fully functional. TNTlite differs from TNTmips only insofar as it limits images to a maximum size of 314 368 cells (e.g. 512 614, 1024 307 pixels), vector objects to 500 points, 1500 lines and 500 polygons, CAD objects to 500 elements, TIN objects to 1500 nodes and database objects to 1500 records per table, and has disabled export capabilities. These limits, however, do not prevent results being printed, and TNTlite can be used for small research projects.

The inclusion of TNTlite in *Image Interpretation in Geology* does not imply any endorsement by either the publishers or myself.

On the CD are 11 Exercises linked to *Image Interpretation in Geology*, which are designed to allow users to master the skills associated with TNTlite processes. These skills are essential in all aspects of geological remote sensing, image interpretation and the production of digital maps to a publishable, professional standard. Each exercise uses data files specially prepared to be compatible with TNTlite. The CD also contains MicroImages' own tutorials and supporting data files, as a series of 54 *Getting Started* booklets in Adobe Acrobat format (the latest version of Acrobat Reader is on the CD too). Those resources cover aspects of remote-sensing and other geospatial applications (land use, urban and environmental) that will be broadly useful to Earth scientists. The full TNTmips *Reference Manual*, which provides more detailed guidance, is also on the CD-ROM.

Randy Smith, PhD of MicroImages Inc and Margaret Andrews of the Open University helped enormously in getting the Exercises working, technically and as teaching material.

Two other kinds of resource reside in separate folders within the *Image Interpretation in Geology* directory (**Imint**) on the CD. One folder contains anaglyph versions of all the stereopair figures in Chapter 4. You can

view them while reading Chapter 4 by using the red-blue spectacles packed with the CD, either in a graphics package, such as Adobe PhotoShop or MS Photo Editor, or in TNTlite after importing these JPEG files to MicroImages format (see later). Other folders contain a collection of different kinds of images, which illustrate various geological themes and types of imagery, and have appended brief descriptions. You should view these using a Web browser because the images are linked to text using HTML—i.e. the compressed images are in a separate sub-folder and therefore accessible by themselves. Many of the images in the collection have sizes that are compatible with TNTlite, so they can be imported as a means of adding interactive, digital interpretation to Exercise 9. They are 24-bit RGB JPEGs and so are limited to three channels. That somewhat limits their use with the other exercises, but the JPEGs can be unpacked to three 8-bit components within TNTlite, for experiments with contrast stretching, some ratioing and filtering.

Contents of the CD-ROM

The CD contains several folders or directories, most of which relate to TNTlite.

TNTlite resources

There are many folders associated with the TNTlite resources on the CD, some of which are copied automatically when you install TNTlite on your hard disk, and others that you have the option to install or use from the CD. The most relevant of those which you can choose to use from the CD are:

Data contains data compatible with TNTlite (in **Data\Litedata**) that are formatted as MicroImages project files (*.rvc), each of which contain several data objects of various kinds. MicroImages' own tutorial exercises (documented in the *Getting Started* Acrobat-format *.pdf files) use them.

Getstart contains illustrated Adobe Acrobat documents (*.pdf files) for each of the MicroImages *Getting Started* tutorials. Folder **Acroread** contains the latest Adobe Acrobat Reader software that automatically runs from the CD when you open an Acrobat document from Getstart by double-clicking on a *.pdf file. If you opted to install the *Getting Started* documents on your hard disk along with TNTlite (see Installing TNTlite below) you can access them from the Help option associated with

each TNTlite process window as you are working, or at any time from the main TNTlite menu (**Display/Getting Started**).

Refman contains the full, 2500 page TNTmips *Reference Manual* as HTML files and linked GIF images (in subfolders **Ci** and **Li**). You can start this from Netscape or Internet Explorer by using **File/Open Page** or **File/Open**, respectively, selecting the CD (e.g. **D:**), then **D:/Refman/Html** and double clicking on **Toc.htm** (table of contents). If you opted to install the *Reference Manual* on your hard disk along with TNTlite (see Installing TNTlite below) you can access relevant parts of the manual from the **Help** option associated with each TNTlite process window as you are working, or at any time from the main TNTlite menu (**Display/Reference Manual**). The manual is comprehensive, and you may find it easier first to refer to the *Getting Started* document appropriate to a process which you find difficult, and going to the *Reference Manual* if you need details about how a process works.

If you have plenty of space on your hard disk, you will find it simplest to install all these folders, as well as the TNTlite software, on your hard disk (see Installing TNTlite, below).

Image Interpretation in Geology resources

The folder **Imint** is where you will find all the material explicitly connected with *Image Interpretation in Geology*. Within it are three folders:

Exercise This contains folders for the text and the data that you will use in the TNTlite geological exercises.

Anagly1 This contains all the anaglyph versions of the stereopairs in Chapter 4. You can easily access them from the CD using either TNTlite or a graphics package, but copy the folder onto your hard disk if you have plenty of space.

Anagly2 and **Imagery**. These contain selections of instructive images from various sources and linked HTML text—more anaglyphs and different types of images, respectively. Again, you can opt either to access them from the CD or copy the folder onto your hard disk.

If you have plenty of space on your hard disk, you will find it simplest to install all these folders on your hard disk (see Installing Resources, below).

Installing resources

Minimum requirements

TNTlite

Hard-disk space: At least 80 Mb for the TNTlite software. *Getting Started* and *Reference Manual* folders take up 60 and 35 Mb, respectively, plus an additional 150

Mb if you wish to install all the *Getting Started* data sets, but you can access them from the CD if your hard-disk space is limited.

CD-drive: Obviously!

Processor: PCs—Pentium III, Pentium II, Pentium Pro, Pentium or 486 computer (or compatible); **MacIntosh**—Power Mac.

Operating system: PCs—Windows 95/98/2000; **MacIntosh**—MacOS 8.0 (version 8.6 or higher is recommended); LINUX.

Random access memory (RAM) At least 16 Mb (32 Mb recommended, especially for MacIntosh). For Macs, **virtual memory** must be set to at least twice that of RAM.

Display: Minimum area of 640 × 480 pixels displaying 256 colours (recommended minimum for *Image Interpretation in Geology* exercises is 1024 × 768 pixels and 16-bit colour).

HTML browser: Netscape Communicator or Internet Explorer.

Adobe Acrobat Reader (the latest version is in **D:/Acroread/Win/Win95nt**).

IIG Resources

Sufficient hard-disk space for the data files associated with the *Image Interpretation in Geology* Exercises. They take up about 140 Mb (including about 100 Mb for the hyperspectral data) and text files for the Exercises (less than one megabyte). If you are short of space, then you can access the folder (**Imint/Exercise/Data/Cuprite**) containing hyperspectral data from the CD-ROM. You must in any event load all the other data associated with the Exercises (**Area1.rvc** etc. in **Imint/Exercise/Data**) onto your hard disk, because you will be saving results into these files. Images and HTML files that illustrate the text of *Image Interpretation in Geology* occupy about 30 Mb.

A graphics package, such as Adobe PhotoShop or MS Photo Editor;

A Web browser, such as Internet Explorer or Netscape Communicator (as high a version as you can get), as all the Exercises are in HTML. If you do not have a browser, or the HTML versions do not run properly, you will also find copies of the exercises in Rich Text Format (RTF, *.rtf) which should be readable in their original format using any word processor less than 10 years old (NOT Notepad or other text-only packages).

Installation

PC users: autorun

If your system supports Autorun CDs, a few seconds after loading the CD a “splash” window will appear on screen. This has 6 buttons:

IIG Menu...
 Install IIG Resources...
 Install TNTlite...
 Run TNTlite...
 Browse...
 Exit

IIG Menu Clicking on this button loads an HTML menu page into your default browser. This provides links to all the IIG Resources on the CD (in folder **Imint**), and allows you to use the resources—including the Exercises and data that relate to the TNTlite geospatial analysis software—directly from the CD.

Install IIG Resources Clicking on this button immediately copies all the IIG resources in the folder **D:/Imint** from the CD to your hard drive, using the DOS utility **xcopy**. The path to the new folder containing the IIG resources will be **C:/Imint**, i.e. it will be a top-level folder on your hard disk.

Install TNTlite Clicking on this button starts the TNTlite installation procedure. You will first have to confirm that you wish to **Continue** with installation (ignore the instruction to attach the software licence key). There now follow some informative windows covering system requirements that you have already read in the previous section. The next important window deals with **Setup Options**, of which you can select only one at a time. Select **Install/Setup the TNTlite version of TNTmips** and press **Continue**. Setup offers you the option of taking the default destination for TNTlite—Path: **C:\Tnt_win**—or choosing another. **Accept the default**, when Setup installs the software and shows you progress. When done, the installation procedure returns to Setup Options window. Click in the window once, and you can decide whether or not to install the *Reference Manual*, the *Getting Started* booklets and the Sample Data associated with them.

Run TNTlite Clicking on this button launches the TNTlite software, provided that you have already installed it in **C:/Tnt_win**. If you have not done this, nothing happens (TNTlite will not run from the CD). The button is provided so that you can use resources and data on the CD if you chose not to copy them to your hard disk because of space problems.

Browse Clicking on this button opens a window showing the contents of the CD. By double clicking on a particular folder, you will see a new window showing its contents and so on. You can open RTF, HTML and PDF files from here, by double-clicking on their icons, which will launch your default word processor, browser or Adobe Acrobat Reader. This is particularly useful for the TNTlite *Getting Started* booklets (in folder **D:/Getstart**), the TNTlite *Reference Manual* (in folder **D:/Refman/html**) or the IIG Resources (various folders in **D:/Imint**), if you have decided not to copy them to your hard disk.

Note: Each time that you need to use the “splash” buttons, you should eject the CD tray and then re-insert it.

PC users: manual

If for some reason Autorun does not display the “splash”, you will need to perform the above manually, using Windows Explorer, as follows:

IIG Menu Display the contents of **D:/Imint**, and double click on **Menu.htm**.

Install IIG Resources Copy **D:/Imint** and all its contents to your hard drive **C:**.

Install TNTlite Double click on **Setup.exe** in **D:**, or use **Start/Run** with the command line **D:\setup.exe**.

Note: Unlike many software packages, installing TNTlite does not create strangely named files and stealthily place them in some other hard-disk folder, where you can never find them. So there is no need to carefully uninstall the software. If you want temporarily to create space for something else, simply delete the whole **C:\TNT_WIN** folder, and re-install TNTlite when you need to.

Run TNTlite Installing TNTlite will automatically add a line to your Start Up menu (**Start/Programs**) from which you can run TNTlite 6.4. But see Tips below.

Browse You can browse the CD, as above, directly from My Computer or Windows Explorer.

Useful tips

Using the Resources can be made easier if you create Shortcuts to several things.

TNTlite When you have completed the installation you will see a small window (**Free TNTlite products**) with several icons. Left-hold-drag the icon labelled **TNTmips 6.4** onto your desktop. Double-clicking on the shortcut will launch TNTlite.

If you have installed the *Reference Manual* on your hard disk, you can create a shortcut to the Table of Contents document on your desktop, so that you can study the manual without having to start TNTlite. Go to **C:/Tnt_win/Refman/Html** in Windows Explorer and right-click on **Toc.htm**. Select **Create Shortcut** from the pull-down menu, and then click-hold-drag the shortcut file onto your desktop from Windows Explorer. Double-clicking on the shortcut will open the Table of Contents in your default browser, and you can select links to all the features of TNTlite.

IIG Resources If you have copied all the Resources to your hard disk, go to **C:/Imint** and right click on **Menu.htm**. Select **Create Shortcut** from the pull-down menu, and then click-hold-drag the shortcut file onto your desktop from Windows Explorer. Double-clicking on the shortcut will open the IIG menu in your default

browser, and you can select links to all the resources associated with the book.

PowerMac users

On a PowerMac Autorun will not function. Instead you will see the contents of the CD in a window when you open the CD icon from the desktop.

Install TNTlite Double click on the **Installer** icon. The next window allows you to select the **License Options**, so select **Install TNTlite ...** and press **OK**. Check **TNTmips 6.4 for PowerMac**, (if you wish to install the *Reference Manual* check the **Reference Manual HTML** option) then press **Install**. TNTlite now loads into a folder on the hard drive automatically named **TNT Products 6.4**.

Note: Unlike many software packages, installing TNTlite does not create strangely named files and stealthily place them in some other hard-disk folder, where you can never find them. So there is no need to carefully uninstall the software. If you want temporarily to create space for something else, simply delete the whole **TNT Products 6.4** folder, and re-install TNTlite when you need to. Because files copied from the CD may be locked, you will need to hold down the **Alt** key when selecting the **Empty Wastebasket** option in order to completely delete them from the hard drive. This also applies if you wish to empty the **Imint** folder from the wastebasket (see below).

Run TNTlite Open the **TNT Products** folder and double click on the **TNTmips** icon. TNT displays a MicroImages logo screen and then opens a full screen window named **MicroImages X Window System (MI/X)**. All TNT processes take place within this unique X Server. Treat the Micro-Images X Window System as a simple background window. Any time during a TNT session, you can use the normal Macintosh techniques (such as the **Apple** menu or the **Application** menu) to jump from MI/X back to the Mac desktop. TNTlite should appear thereafter as a **recently used application** in the **Apple** pull-down.

Install IIG Resources Copy the **Imint** folder to your hard disk. For ease of access to the IIG Resources you will need to launch **Menu.htm** from this folder. It will appear as a **recently opened** document in the **Apple** pull-down menu thereafter.

Working with the IIG resources

The whole package of resources that relate to *Image Interpretation in Geology* is designed for access using a browser (Netscape or Internet Explorer) from the HTML file **Menu.htm** in the **Imint** folder on your hard disk (or as a desktop shortcut on a PC), which links to all

the resources, except for TNTlite. You need to launch TNTlite separately, either from **Start/Programs** or a desktop shortcut (PC), or using the **recent applications** and **documents** in the **Apple** menu (PowerMac).

TNTlite uses MI/X, MicroImages' X-server that operates within Windows or MacOS (it is similar in some respects to the UNIX windowing system). The X window that opens, and within which all TNTlite windows appear, when you launch TNTlite, has one problem—you cannot easily resize it. That does not mean that TNTmips and other Windows or Mac software will not work at the same time. The X Window merely hides them. That is no great problem, as both PCs and PowerMacs have means whereby you can bring other windows to the front of the display—the **Auto Hide** and **On Top** functions for the Windows and MS Office Task Bars, in the case of a PC. However, every time you click inside the X Window the others hide again. You will need to find ways to shift from one set of windows to the other. A crude option is to minimize the X Window—the **'** box at top right, but it is better to bring Windows or MacOS windows to the front when you need to, using the Task Bar in the first case and toggling the X-window using the **Finder** Icon on a Mac. So that you can view both the text for the Exercises and TNTlite windows, the larger the display area for which you have set your monitor, the more you will be able to see when reading instructions. An area of 1024 × 768 pixels is OK, but 1280 × 1024 is obviously better.

Once you begin the IIG Exercises, resize the window that you use to view the text to reveal as much as possible of the TNTlite windows. HTML files automatically fit the width of a browser window, so will pose no problems. If you view the RTF format text files for the Exercises in a word-processor window, you may find that some of the text is hidden, depending on your software. Using the **View/Online Layout** option in recent versions of MS Word automatically wraps text to the width that you choose. If this is not possible, then you will need to reset the line length for the text files to the width of your resized window.

Getting help

Like all image processing and desk-top mapping packages, TNTlite is complex and completing the exercises involves your mastering many different windows. It is possible to get stuck or lost! The *Getting Started* booklets are very useful in resolving common problems, as is the *Reference Manual*. Two booklets entitled *Getting Started: Displaying Geospatial Data* and *Getting Started: Navigating* are particularly useful in helping you learn to navigate around in TNTlite and display your results. I have tried to follow MicroImages' style in the IIG exercises, but to

save space they do not repeat the basic TNTlite X-Windows operations. It is essential that you begin with Exercise 1, which explains the most important, basic aspects of that style of operation. There are reminders in later exercises of these basics, but not often, so as to build up your skills in as short a space as possible. So, you can resolve any difficulties as you mount the learning 'ramp' if you return briefly to the earlier exercises to refresh your memory.

If you find any parts of the exercises difficult to follow, even after looking for general guidance in the *Getting Started* booklets, please contact me via Blackwell Science (but do persevere for longer than a few moments before writing). That will also help me to refine the instructions for a revised version of the CD. I will NOT answer questions relating to operating systems or hardware, expecting you to be thoroughly familiar with your own computer or having direct access to someone who is.

MicroImages is a commercial company, with primary duty to those clients who have purchased the pro-

fessional version of TNTmips. MicroImages software-support engineers give preference to those clients. However, as time allows, they will respond sympathetically to queries from TNTlite users. Use support@microimages.com for e-mail: and #1-402-477-9559 for fax enquiries (please do not phone) to contact MicroImages Software Support.

MicroImages continually develops its products, generally releasing new versions on a quarterly basis. Anyone can obtain upgrades for TNTlite either by downloading from the Web (<http://www.microimages.com>) or requesting a CD (MicroImages may impose a small charge to cover shipping the CD) from:

MicroImages, Inc.

11th Floor—LBL Plaza South

206 South 13th Street

Lincoln, NE 68508-2010

USA

Fax: #1-402-477-9559

e-mail: info@microimages.com

Appendix D | Sources of Image Data

Image atlases

These may be quite dated references, before the use of CD and Web image resources, but they constitute a timeless source of useful images, particularly suited for geological interpretation. For individual use, it is possible to use a desktop scanner to import your selection of images, making sure that the remove screens option is set because all use photolitho printing.

- Ford, J.P., Blom, R.G., Bryan, M.G., Daily, M.I., Dixon, T.H., Elachi, C. & Xenos, E.C. (1980) *Seasat Views North America, the Caribbean, and Western Europe with Imaging Radar*. Publication 80-67, Jet Propulsion Laboratory, Pasadena, CA.
- Ford, J.P., Cimino, J.B. & Elachi, C. (1983) *Space Shuttle Columbia Views the World with Imaging Radar—the SIR-A Experiment*. Publication 82-95, Jet Propulsion Laboratory, Pasadena, CA.
- Hamblin, W.K. (1980) *Atlas of Stereoscopic Aerial Photographs and Landsat Imagery of North America*. Tasa, Minneapolis, Minnesota.
- Institut Geographique National (1984) *l'Atlas Des Formes Du Relief*. Institut Geographique National, Paris.
- Mollard, J.D. (1974) *Landforms and Surface Materials of Canada: a Stereoscopic Airphoto Atlas and Glossary*. J.D. Mollard, Regina, Saskatchewan.
- Sheffield, C. (1981) *Earthwatch: a Survey of the World from Space*. Sidgwick and Jackson, London.
- Sheffield, C. (1983) *Man on Earth*. Sidgwick and Jackson, London.
- Short, N.M. & Blair, R.W., Jr (1986) *Geomorphology from Space*. NASA SP-486, US Government Printing Office, Washington, DC.
- Short, N.M., Lowman, P.D., Freden, S.C. & Finch, W.A. (1976) *Mission to Earth: Landsat Views the World*. NASA Special Publication 360. NASA, Washington DC.
- Short, N.M. & Stuart, L.M. (1982) *The Heat Capacity Mapping Mission (HCMM) anthology*. NASA Special Publication 465. NASA, Washington DC.

Websites

These URLs lead you to pages either where digital images can be viewed and downloaded, or where you can find useful items on remote sensing and related subjects. Many of the sites have lists of useful links, so that you can build up a comprehensive list of bookmarks yourself.

Tips: Most images displayed through an Internet browser will end up in your temporary cache, from where you can copy them to permanent storage, but they will often have incomprehensible names. A better option is to right-click on the displayed image, select Save

Image As: from the drop-down menu, and then assign your own name and a suitable folder before saving. In most cases the files will be either JPEGs or GIFs, with the appropriate extension. However, some sites that allow you to browse global archives of data as reduced resolution images assign odd or nonexistent extensions in the hope that no-one will think of doing this. They are usually JPEGs, so assign the extension .jpg when saving. Also, remember all modern computers have a Print Screen button that saves the screen to RAM. Try using Alt+Print Screen, which copies the active window to RAM, and then pasting to a graphics package. You can have lots of fun piecing together mosaics of such global browse images, and can find educational or research uses for them. Eritrea.bmp in the Images folder within the IIG directory is such an image.

Note: Some producers of remotely-sensed image data have copyright in them, whereas others waive rights. You should check carefully on the websites for the copyrights held by the data producers or suppliers. If you wish to disseminate downloaded data then you **must** obtain permission from the source.

ASTER HOME PAGE (JPL)

<http://asterweb.jpl.nasa.gov/>

ASTER site, Japan

<http://astweb.ersdac.or.jp/ao/>

British Geological Survey—Geoscience Information Resources

<http://www.bgs.ac.uk/bgs/relat.html>

Digital Chart of the World

<http://www.maproom.psu.edu/dcw/>

EarthWatch

<http://www.digitalglobe.com/>

Environmental Research Institute of Michigan

<http://www.erim-int.com/>

EOS Homepage

<http://eosps0.gsfc.nasa.gov/>

Eurimage S.p.A.

<http://www.eurimage.com/>

Goddard Space Flight Center

http://www-v0ims.gsfc.nasa.gov/v0ims/DATAPRODSERV/Data_Prod_Types_Serv.html

Indian Remote Sensing Satellites

<http://202.54.32.164/>

Landsat 7

<http://landsat7.usgs.gov/>

Marine Geology & Geophysics Images

<http://www.ngdc.noaa.gov/mgg/image/images.html>

Marine Terraserver

<http://www.ngdc.noaa.gov/mgg/image/images.html>

MicroImages Atlases

<http://microimages.com/atlasserver>

Microsoft Terraserver

<http://terraserver.microsoft.com/default.asp>

NASA Earth Science Enterprise

<http://www.earth.nasa.gov/>

NASA Jet Propulsion Laboratory

<http://www.jpl.nasa.gov/>

NOAA Home Page

<http://www.noaa.gov/>

RADARSAT

<http://radarsat.space.gc.ca/>

Remote Sensing Links

<http://www.cent.org/geo12/focus5e.htm> and

<http://www.ersi.bc.ca/links.html>

SeaWiFS Project—Homepage

<http://seawifs.gsfc.nasa.gov/SEAWIFS.html>

Shuttle Radar Topography Mission (SRTM)

<http://www-radar.jpl.nasa.gov/srtm/>

SIR-C Precision Data

<http://edcwww.cr.usgs.gov/landdaac/sir-c/precision/precision.html>

Space Imaging—IKONOS

<http://www.spaceimagingeurope.com/>

Spot Image

<http://www.spotimage.fr/>

TERRA (EOS AM-1)

<http://terra.nasa.gov/>

TOPEX/Poseidon

<http://topex-http://www.jpl.nasa.gov/science/science.html>

USGS Global Land Information System (GLIS)

<http://edcwww.cr.usgs.gov/webglis/>

USGS Spectral Laboratory

<http://speclab.cr.usgs.gov/>

Virtual Field Course

http://www.geog.le.ac.uk/vfc/index_2.html

WWW Virtual Library: Remote Sensing

<http://www.vtt.fi/aut/rs/virtual/>

X-SAR data at DLR, Germany

<http://www.op.dlr.de/ne-hf/SRL.html>

Glossary

This Glossary defines many terms and concepts with which readers may not be familiar and which crop up in the book and in other remote-sensing literature. It should be used in conjunction with the Index, which shows where the terms are defined more fully and in context.

absolute temperature Temperature measured on the Kelvin scale, the base of which is absolute zero (-273°C). 0°C is expressed as 273 K.

absorptance A measure of the ability of a material to absorb electromagnetic (EM) energy at a specific wavelength.

absorption band A range of wavelengths in the electromagnetic (EM) spectrum where a material absorbs EM energy incident upon it, often resulting from various energy-matter transitions.

achromatic line A line in red, green, blue image display space, running at 45° to the axes. A three-channel combination from most images tends to be correlated, and so the multispectral data will plot close to this line, and the image will not be strongly coloured unless special image processing techniques (such as **decorrelation stretching**) are applied.

achromatic vision The perception by the human eye of changes in brightness, often used to describe the perception of monochrome or black and white scenes.

active remote sensing A system based on the illumination of a scene by artificial radiation and the collection of the reflected energy returned to the system. Examples are **radar** and systems using lasers.

acuity A measure of human ability to perceive spatial variations in a scene. It varies with the spatial frequency, shape and contrast of the variations, and depends on whether the scene is coloured or monochrome.

additive primary colours The spectral colours red, green and blue, which Thomas Young (1773–1829) discovered to be capable of reproducing all other colours when mixed by projection through filters, and each of which cannot be produced by mixtures of the other two.

AIS Airborne Imaging Spectrometer.

albedo The fraction of the total electromagnetic (EM) energy incident on a material that is reflected in all directions.

analogue image An image where the continuous variation in the property being sensed is represented by a continuous variation in image tone. In a photograph this is achieved directly by the grains of photosensitive chemicals in the film, in an electronic scanner, the response in say millivolts is transformed to a display on a cathode-ray tube where it may be photographed.

artefact A feature on an image that is produced by the optics of the system or by digital image processing, and sometimes masquerades as a real feature.

ASTER Advanced Spaceborne Thermal Emission and Reflection Radiometer to be carried by EOS-1AM (Terra), launched in December 1999. ASTER will acquire images in 14 spectral bands in the reflected and thermally emitted regions, selected particularly for geological applications.

atmospheric shimmer An effect produced by the movement of masses of air with different refractive indices, which is seen most easily in the twinkling of stars. Shimmer results in blurring on remotely sensed images, and is the ultimate control over the resolution of any system.

atmospheric window A range of electromagnetic (EM) wavelengths where radiation can pass through the Earth's atmosphere with relatively little attenuation.

attribute Property assigned to a line, polygon or point in vector-format data. Several may be combined to give a description of a vector that is useful in a geographical information system (GIS).

AVHRR Advanced Very High Resolution Radiometer, a multispectral imaging system carried by the TIROS-NOAA series of meteorological satellites.

AVIRIS Airborne Visible/Infrared Imaging Spectrometer.

azimuth In general this is the compass bearing of a line given in degrees clockwise from north. In **radar** terminology it refers to the direction at right angles to the radar propagation direction, which is parallel to the ground track in a sideways-looking system.

band In remote sensing, a range of wavelengths from which data are gathered by a recording device.

band-pass filter Spatial filter designed to accentuate only features with frequencies between set limits.

band-stop filter Spatial filter designed to selectively remove features with frequencies between set limits.

bin One of a series of equal intervals in a range of data, most commonly employed to describe the divisions in a histogram.

binary A numerical system using the base 2. Examples are $0 = 0$, $1 = 2^0 = 1$, $10 = 2^1 = 2$, $11 = 2^1 + 2^0 = 3$.

bit An abbreviation of binary digit, which refers to an exponent of 2. A bit is represented by 0 or 1 for 'on' or 'off' in a digital computer.

blackbody A perfect radiator and absorber of electromagnetic (EM) energy, where all incident energy is absorbed and the energy radiated from the body at a particular temperature is at the maximum possible

- rate for each wavelength, as governed by the **Stefan-Boltzmann Law**. No natural material has these ideal properties, although water is a close approximation.
- blind spot** The point of entry of the optic nerve to the retina where no radiation is detected by the eye.
- bond-stretching and bond-bending transitions** Produce spectral features due to distortion of bonds in molecules, particularly the metal-hydroxyl bonds Al-OH and Mg-OH.
- box classification** A classification technique in which the decision boundary around the training set is a rectangular box. The term **parallelepiped classification** is synonymous.
- box filter** Rectangular array of numbers used in **convolution** with an image to implement some kind of **spatial-frequency filtering**.
- brightness** Often used loosely to refer either to the amount of radiation coming from a surface or to the appearance of a region within an image.
- byte** A group of 8 bits of digital data in **binary** form. A byte represents digital numbers up to 255, and is the standard adopted by most digital remote-sensing systems where the range of energies is coded from 0 to 255.
- canonical analysis** A variant of **principal component analysis** that uses a selected range of the input variables, usually relating to specific surface materials, to generate the statistics used in the selection of principal component axes.
- cell assemblies** The linked receptors, retinal neurons and neural cells in the visual cortex of the brain that enable interaction between perception and past experience.
- change detection** A range of techniques to identify important differences between scenes recorded on different dates.
- channel** The range of wavelengths recorded by a single detector to form an image. A multispectral image is recorded in several channels simultaneously. The term also can be used to refer to a synthetic channel (i.e. not a simple spectral band as originally recorded), such as one created by ratioing or principal components transformation. See also **band**.
- charge transfer** A mechanism of spectral absorption of EMR, when an electron is transferred across a chemical bond.
- charge-coupled device (CCD)** A light-sensitive capacitor whose charge is proportional to the intensity of illumination. They are able to be charged and discharged very quickly, and are used in **push-broom** devices, **spectroradiometers** and modern video cameras.
- chromatic vision** The perception by the human eye of changes in hue.
- CIE colour system** **Tristimulus colour theory** based on mathematically derived chromaticity co-ordinates, produced by the **Commission Internationale de l'Éclairage**.
- classification** The process of assigning individual pixels of a multispectral image to categories, generally based on spectral characteristics of known parts of a scene.
- Coastal Zone Colour Scanner (CZCS)** A multispectral imaging system carried by the **Nimbus** series of meteorological satellites.
- coherent radiation** Electromagnetic radiation whose waves are equal in length and are in phase, so that waves at different points in space act in unison, as in a **laser** and **synthetic aperture radar**.
- cones** Receptors in the retina that are sensitive to colour. There are cones sensitive to red, green and blue components of light.
- context** The known environment of a particular feature on an image.
- contrast** The ratio between the energy emitted or reflected by an object and its immediate surroundings.
- contrast stretching** Expanding a measured range of **digital numbers (DN)** in an image to a larger range, to improve the contrast of the image and its component parts.
- convolution** Arithmetic combination between two arrays of numbers, generally implemented by a small **box filter** whose weights multiply with DN in the array overlain by it, the sum of the pixel-by-pixel operations being substituted for the pixel at the centre of the area overlain by the box. The convolved image is generated by systematically passing the box filter over every pixel in the scene.
- corner reflector** A cavity formed by three planar reflective surfaces intersecting at right angles, which returns **radar** directly back to its source.
- crystal-field effect** Dependence of spectral features on co-ordination of ions in a mineral molecule. Most commonly shown by minerals dominated by ferric iron (Fe^{3+}).
- cut off** The digital number in the histogram of a digital image that is set to zero during **contrast stretching**. Usually this is a value below which **atmospheric scattering** makes a major contribution.
- dark-pixel correction** Use of the darkest pixels in a scene, either total shadows or deep, clear water, to approximate the contribution by **atmospheric scattering**.
- decorrelation stretching** A way of making a three band multispectral image more colourful by stretching the data distribution along axes related to the natural elongation of the data distribution (the **principal component** directions). After stretching, the data are displayed on the original red, green and blue axes, thus retaining the original, straightforward colour relationships. Sometimes abbreviated to D-stretch.
- density slicing** The process of converting the full range of data into a series of intervals or slices, each of which expresses a range in the data.

depression angle In radar usage this is the angle between the horizontal plane passing through the antenna and the line connecting the antenna to the target. It is easily confused with the look angle.

depth of field The range of distances in focus using a lens.

diffuse reflection Reflection of light (or other electromagnetic radiation) from a surface approximately evenly in all directions, as opposed to specular (mirror-like) reflection. A diffuse reflector has a surface that is rough on the scale of the wavelength of electromagnetic radiation concerned.

digital image An image where the property being measured has been converted from a continuous range of analogue values to a range expressed by a finite number of integers, usually recorded as binary codes from 0 to 255, or as one byte.

digital number (DN) The value of a variable recorded for each pixel in an image as a binary integer, usually in the range 0–255. An alternative term used in some texts is brightness value (BV). The plural of its acronym is DN.

directional filter A spatial-frequency filter that enhances features in an image in selected directions.

DLR German Space Agency.

Doppler shift A change in the observed frequency of electromagnetic (EM) or other waves caused by the relative motion between source and detector. Used principally in the generation of synthetic-aperture radar images.

edge A boundary in an image between areas with different tones.

edge enhancement The process of increasing the contrast between adjacent areas with different tones on an image.

eigenvalue Proportion of the variance in multispectral data that is redistributed to a principal component.

eigenvector Crudely, the contribution of an input band to a principal component.

electromagnetic radiation (EMR) Energy transported by the propagation of disturbances in the electric and magnetic fields. Detection and measurement of natural or artificial EMR is the basis for remote sensing.

electronic transition A mechanism of spectral absorption of electromagnetic radiation (EMR), when an electron is moved from a lower orbit to a higher, as the result of absorption of a photon. Also a mechanism of generation of EMR, when an electron goes from a higher orbit to a lower, and emits a photon in the process.

emissivity A measure of how well a surface emits, or radiates, electromagnetic radiation (EMR) thermally; defined as the ratio between the thermal exitance from the surface and the thermal exitance from a black body (perfect emitter) at the same temperature. A blackbody therefore has an emissivity of 1 and natural materials range from 0 to 1.

emittance A term for the radiant flux of energy per unit area emitted by a body. Now obsolete.

emitted region Part of the spectrum between 3 μm and 1 cm dominated by energy emitted by bodies above absolute zero.

EOS Earth Observing System, a proposed multinational series of heavily instrumented remote sensing satellites to be deployed in the decade beginning in the mid-1990s. It probably will include American (NASA), European (ESA) and Japanese polar orbiting platforms and a manned space station.

EOSAT Earth Observation Satellite Company, based in Lanham, Maryland, USA. A private company contracted to the US Government since September 1985 to market Landsat data and develop replacements for the Landsat system.

equalization stretch Contrast stretch that forces the cumulative histogram of DN to a straight line, thereby spacing bins according to their frequencies. This achieves an equal population density of pixels along the DN axis. Also known as **equipopulation** or **ramp stretch**.

equatorial orbit An orbit of a satellite around the Earth in which the orbital plane makes an angle of less than 45° with the Equator.

equipopulation stretch See **equalization stretch**.

EROS Earth Resources Observation System, based at the EROS Data Center, Sioux Falls, South Dakota, USA. Administered by the US Geological Survey, it forms an important source of image data from Landsat-1, -2 and -3 and for Landsat-4 and -5 data up to September 1985, as well as airborne data for the USA.

ERS-1 Earth Remote Sensing Satellite-1, a European (ESA) satellite scheduled for launch in 1992, targeted principally on marine applications by the use of active and passive microwave techniques.

ESA European Space Agency, based in Paris. A consortium between several European states for the development of space science, including the launch of remote-sensing satellites.

exitance The radiant flux density (of electromagnetic radiation) leaving a surface.

exponential stretch Contrast stretch using a **look-up table** that is an exponential curve. Accentuates contrast in light regions.

false-colour image A colour image where parts of the non-visible electromagnetic (EM) spectrum are expressed as one or more of the red, green and blue components, so that the colours produced by the Earth's surface do not correspond to normal visual experience. Also called a false-colour composite (FCC). The most commonly seen false-colour images display the very-near-infrared region as red, red as green and green as blue.

flat iron A term used to describe roughly triangular features seen on images, derived from their similarity to the household appliance used for smoothing cloth. On aerial photographs they result from dipping rock layers. On **radar** images they may be an artefact caused by high relief and steep **depression angle**.

flat-field calibration Use of spectrally featureless materials, such as halite, to judge atmospheric effects in **hyper-spectral data**. The observed spectrum of spectrally flat material is used to divide those of other pixels, thereby making them roughly equivalent to reflectance spectra determined under laboratory conditions.

fluorescence A property of some materials where electromagnetic (EM) energy of one wavelength is absorbed and then re-emitted at a longer wavelength.

foreshortening A distortion in **radar** images causing the lengths of slopes facing the antenna to appear shorter on the image than on the ground. It is produced when radar wavefronts are steeper than the topographic slope.

Fourier synthesis and analysis Construction and disassembly, respectively, of a complex waveform in terms of sine waves with different wavelengths and amplitudes. Used in frequency-domain filtering of images.

fovea The region around that point on the retina intersected by the eye's optic axis, where receptors are most densely packed. It is the most sensitive part of the retina.

frequency The number of waves that pass a reference point in unit time, usually 1 s.

f-stop Ratio between focal length and aperture diameter of a lens.

fundamental Refers to the main spectral feature produced by a **vibrational transition**.

gamma-ray spectrometer Based on light emission by a crystal when exposed to gamma radiation. By tuning the system to the energy levels of gamma radiation emitted by ^{40}K , ^{232}Th and ^{238}U , it discriminates between and measures the radiation emitted by potassium, thorium and uranium contained by surface materials.

Gaussian stretch See **normalized stretch**.

geographical information system (GIS) A data handling and analysis system based on sets of data distributed spatially in two dimensions. The data sets may be map orientated, when they comprise qualitative attributes of an area recorded as lines, points and areas in **vector format**, or image orientated, when the data are quantitative attributes referring to cells in a rectangular grid in **raster format**.

georeferencing Correlation of positions on images with geographic co-ordinates, using ground-control points. Sometimes referred to as geocoding.

Geoscan Remote-sensing device operated by Carr-Boyd Pty of Australia, which incorporates narrow-band spectral coverage from both the reflected and emitted regions.

geostationary orbit An orbit at 41 000 km in the direction of the Earth's rotation, which matches speed so that a satellite remains over a fixed point on the Earth's surface.

greybody A material with constant emissivity less than 1.0 for all wavelengths.

grid format The result of interpolation from values of a variable measured at irregularly distributed points, or along survey lines, to values referring to square cells in a rectangular array. It forms a step in the process of contouring data, but also can be used as the basis for a **raster format** to be displayed and analysed digitally after the values have been rescaled to the 0–255 range.

ground-control point A point in two dimensions that is common to both an image and a topographic map, and can be represented by (*x,y*) co-ordinates based on the map's cartographic projection and grid system. Used in geometric correction of distorted images, and their registration to a convenient map projection.

ground range Distance on the Earth's surface from the position of an object to that directly below a **radar** antenna.

ground truth Observations made on the ground that can be used to verify the interpretation of an image. Ground truth includes field-based identification of surface types and their distribution (to help perform and verify classification). Comparable observations at sea (e.g. phytoplankton concentration, wave height, etc.) are usually known as sea truth.

harmonic Secondary spectral feature related to vibrational transitions at wavelengths related to that of the **fundamental** (analogous to those in sound). Also known as an **overtone**.

heat capacity A measure of how much energy it takes to warm up a given mass of a material through 1°C.

Heat Capacity Mapping Mission (HCMM) NASA satellite launched in 1978 to observe thermal properties of rocks and soils. Remained in orbit for only a few months.

high-pass filter A spatial filter that selectively enhances contrast variations with high spatial frequencies in an image. It improves the sharpness of images and is a method of edge enhancement.

HIRIS High Resolution Imaging Spectrometer, possibly to be carried by a future platform of the **Earth Observing System**.

histogram A means of expressing the frequency of occurrence of values in a data set within a series of equal ranges or bins, the height of each bin representing the frequency at which values in the data set fall within the chosen range. A cumulative histogram expresses the frequency of all values falling within a bin and lower in the range. A smooth curve derived mathematically from a histogram is termed the **probability density function (PDF)**.

HRV Haute Resolution Visible, the visible and very-near-infrared pushbroom imaging system carried by SPOT; 20 m pixels in colour and 10 m pixels in panchromatic mode.

hue One of the three characteristics that define the colour of a pixel in an image in terms of **intensity**, hue and **saturation**. Hue is a measure of the relative amounts of red, green and blue contributing to the colour.

Hyperion Imaging spectrometer with 7.5 km swath width to be carried on long-delayed NASA satellite (EO-1); part of NASA's latest strategy of low-cost imaging systems.

hyperspectral data Often used for multi-channel, narrow band data captured by **imaging spectrometers**.

image dissection The breaking down of a continuous scene into discrete spatial elements, either by the **receptors** on the retina, or in the process of capturing the image artificially.

image striping A defect produced in **line-scanner** and **pushbroom** imaging devices produced by non-uniform response of a single detector, or amongst a bank of detectors. In a line-scan image the stripes are perpendicular to flight direction, but parallel to it in a pushbroom image.

imaging radar A **radar** that constructs an image of the terrain or sea surface by complex processing of the echoes reflected back to the antenna.

imaging spectrometer A sophisticated remote sensing instrument that records an image in a large number of narrow spectral channels.

incidence angle The angle between the surface and an incident ray of electromagnetic (EM) radiation—most usually referring to **radar**.

infrared radiation Electromagnetic radiation (EMR) of wavelength between 0.7 μm and 100 μm , comprising reflected infrared 0.7–3.0 μm and thermal infrared 3.0–100 μm .

instantaneous field of view (IFOV) The solid angle through which a detector is sensitive to radiation. It varies with the intensity of the radiation, the time over which radiation is gathered, and forms one limit to the resolution of an imaging system.

intensity A measure of the energy reflected or emitted by a surface. Specifically, the overall brightness of a colour in the **IHS** colour system.

intensity-hue-saturation (IHS) colour system Based on a cone whose axis has equal values for red, green and blue (the grey axis). **Intensity** of a colour is measured by projection of its RGB co-ordinates on to the grey axis. A colour's **hue** is the angular position (from 0–360°) of its RGB co-ordinates around the circumference of a circular section through the cone, known as a colour wheel. **Saturation** measures the distance of a colour's RGB co-ordinates radially from the grey axis.

intensity-hue-saturation (IHS) processing After converting a colour image from Cartesian RGB colour axes to the conical **IHS** colour system, its intensity, saturation or hue can be modified or substituted. Most commonly, this is used in enhancing the visibility of colours in an image by stretching the **intensity** and **saturation**, without altering the **hue**. This maintains the original relationship between colour and spectral properties.

interferometric radar (InSAR) See **radar interferometry**.

irradiance The radiant flux density falling on a surface.

IRS-1 Indian Remote Sensing satellite launched in 1988, which carries two **pushbroom** scanners covering four bands in the **visible** and **VNIR**, one with 73 m resolution the other creating pixels half that size. Most data are restricted to the Indian subcontinent.

JERS-1 Japanese Earth Resources Satellite, launched in 1992, when it was renamed 'Fuyo-1' (the acronym JERS-1 seems to have stuck outside Japan). JERS-1 carries four sensors. **Pushbroom** sensors in the reflected region produce two bands in the **visible** range, two in the **VNIR**, one being a forward-pointing band to produce stereoscopic images, and four bands in the short-wavelength infrared (**SWIR**). Each band has a pixel size of 18.3 by 24.2 m. The fourth sensor is an L-band SAR system with a ground resolution of 18 m. Data are available globally through onboard recording facilities.

Kirchhoff's Law Simple relationship between **reflectance** and **emissivity** for good emitters of radiation, whose emissivities and absorptivities are equal.

Lambertian reflector A perfectly diffuse reflector, which reflects incident electromagnetic radiation (EMR) equally in all directions.

Landsat A series of remote-sensing satellites in Sun-synchronous polar orbit that began in 1972. Initially administered by NASA, then NOAA, and since 1985 by EOSAT. Carried **MSS**, **RBV** and **TM** systems.

laser Light artificially stimulated electromagnetic radiation. A beam of coherent radiation with a single wavelength.

layover A distortion in **radar** images when the angle of surface slope is greater than that of radar wavefronts. The base of a slope reflects radar after the top of the slope, and because radar images express the distance to the side in terms of time, the top appears closer to the platform than the base on an image, giving the impression of an overhanging slope.

Lewis Satellite carrying a 384-band **imaging spectrometer**, which sadly failed to achieve orbit in 1998.

line drop out The loss of data from a scan line caused by malfunction of one of the detectors in a **line-scanner**.

line-scanner An imaging device that uses a mirror to sweep the ground surface normal to the flight path of the platform. An image is built up as a strip comprising lines of data.

linear stretch Contrast stretch that redistributes minimum and maximum DN in raw data linearly between 0 and 255 respectively, thereby stretching all intermediate DN over the full range of brightnesses.

logarithmic stretch Contrast stretch using a **look-up table** that is a logarithmic curve. Accentuates contrast in dark regions.

look angle The angle between the vertical plane containing a **radar** antenna and the direction of radar propagation. Complementary to the **depression angle**.

look direction The direction in which pulses of **radar** are transmitted.

look-up table (LUT) A mathematical formula used to convert one distribution of data to another, most conveniently remembered as a conversion graph.

luminance A measure of the luminous intensity of light emitted by a source in a particular direction.

Mach band An optical illusion of dark and light fringes within adjacent areas of contrasted tone. It is a psychophysiological phenomenon that aids human detection of boundaries or edges.

mask A geographical area to be separated from other areas for some purpose, usually to perform some image-processing or **GIS** operation without influencing areas outside the mask. A simple example would be bodies of water on which contrast enhancement is to be performed, either to render all in a single tone or to display subtle differences, without affecting land areas.

maximum-likelihood classification A classification technique in which unknown pixels are assigned to a class according to statistically defined probabilities.

median filter A spatial filter that substitutes the median value of digital number (DN) from surrounding pixels for that recorded at an individual pixel. It is useful for removing random noise.

Meteosat The European (ESA) member of the GOES geostationary weather satellite series.

microwave region That part of the spectrum beyond 1 mm.

microwaves Electromagnetic radiation (EMR) of wavelength between 100 μm and 1 m.

mid-infrared (MIR) The range of electromagnetic (EM) wavelengths from 8 to 14 μm dominated by emission of thermally generated radiation from materials. Also known as **thermal infrared**.

Mie scattering The scattering of electromagnetic (EM) energy by particles in the atmosphere with comparable dimensions to the wavelength involved.

minimum-distance-to-means classification A classification technique in which unknown pixels are assigned to whichever class whose mean they are closest to, in multispectral space.

minus-blue photograph A **panchromatic** black and white photograph from which the blue part of the visible range has been removed using a yellow filter.

mixed pixel A pixel in which the digital number (DN) represents the average energy reflected or emitted by several types of surface present within the area that it represents on the ground. Sometimes called a **mixel**.

modulation transfer function (MTF) A measure of the sensitivity of an imaging system to spatial variations in contrast.

MOMS Modular Optical-electronic Multispectral Scanner, a German imaging system carried experimentally by Shuttle missions, and which is based on a **push-broom** system.

MSS Multispectral Scanner, a remote sensing instrument carried on all the members of the **Landsat** series, recording pixels about 80 m across.

multispectral scanner A line-scanner that simultaneously records image data from a scene in several different wavebands. Most commonly applied to the four-channel system with 80 m resolution carried by the **Landsat** series of satellites.

multispectral space The multidimensional space (sometimes known as feature space) represented mathematically by plotting the digital number (DN) in each spectral band of an image on an orthogonal set of axes. If there are six spectral bands under consideration (for example when using the six reflected bands in **Landsat TM** data), then the multispectral space is six-dimensional.

Munsell colour system Arbitrary division of colours based on hue, value and chroma.

nadir The point on the ground vertically beneath the centre of a remote-sensing system.

NASA National Aeronautics and Space Administration, USA.

NASDA National Space Development Agency, Japan.

near-infrared (NIR) The shorter wavelength range of the infrared region of the electromagnetic (EM) spectrum, from 0.7 to 2.5 μm . It is often divided into the very-near-infrared (VNIR) covering the range accessible to photographic emulsions (0.7–1.0 μm), and the short-wavelength infrared (SWIR) covering the remainder of the NIR atmospheric window from 1.0 to 2.5 μm .

Nimbus-7 An experimental polar orbiting satellite launched in 1978, notable for carrying the Coastal Zone Colour Scanner (CZCS).

NOAA National Oceanic and Atmospheric Administration, USA.

noise Random or regular **artefacts** in data, which degrade their information-bearing quality and are due to defects in the recording device.

nonselective scattering The scattering of electromagnetic (EM) energy by particles in the atmosphere that are much larger than the wavelengths of the energy, and which causes all wavelengths to be scattered equally.

nonspectral hue A hue which is not present in the spectrum of colours produced by the analysis of white light by a prism or diffraction grating. Examples are brown, magenta and pastel shades.

normalized difference vegetation index (NDVI) A vegetation index that calculates how far a pixel plots from the soil line, using the relationship: $NDVI = (VNIR - red) / (VNIR + red)$.

normalized stretch Contrast stretch in which a range of raw DN are transformed by forcing their **histogram** to the shape of a normal or Gaussian distribution that spans the full 0–255 range.

opponents colour theory Representation of colour in terms of three components, white–black, blue–yellow and red–green.

orbital period The time taken by a satellite to complete an orbit.

orthophotograph A vertical aerial photograph from which the distortions due to varying elevation, tilt and surface topography have been removed, so that it represents every object as if viewed directly from above, as in a map.

orthophotoscope An optical-electronic device which converts a normal vertical aerial photograph to an orthophotograph.

overtone See **harmonic**.

panchromatic Covering the range of wavelengths in the visible (and sometimes the VNIR) part of the spectrum.

parallax The apparent change in position of an object relative to another when it is viewed from different positions. It forms the basis of stereopsis.

parallax difference The difference in the distances on overlapping vertical photographs between two points, which represent two locations on the ground with different elevations.

parallelepiped classification A classification technique in which the decision boundary around the training set is a rectangular box. The term **box classification** is synonymous.

passive microwaves Radiation in the 1 mm to 1 m range emitted naturally by all materials above absolute zero.

passive remote sensing The capture of images representing the reflection or emission of electromagnetic (EM) radiation that has a natural source.

pattern A regular assemblage of tone and texture on an image. Often refers to drainage systems.

photographic infrared That part of the near-infrared (or reflected infrared) spectrum to which photographic films respond, i.e. wavelengths between 0.7 μm and about 1.0 μm . Virtually synonymous with **very-near-infrared**.

photon A **quantum** of electromagnetic (EM) energy.

photopic vision Vision under conditions of bright illumination, when both rods and cones are employed.

piece-wise stretch Contrast stretch that assigns several ranges of DN in raw data to different output ranges, each being transformed linearly.

pixel A single sample of data in a digital image, having both a spatial attribute—its position in the image and a usually rectangular dimension on the ground—and a spectral attribute—the intensity of the response in a particular waveband for that position (DN). A contraction of picture element.

Planck's Law An expression for the variation of **emittance** of a blackbody at a particular temperature as a function of wavelength.

point spread function (PSF) The image of a point source of radiation, such as a star, collected by an imaging device. A measure of the spatial fidelity of the device.

polar orbit An orbit that passes close to the poles, thereby enabling a satellite to pass over most of the surface, except the immediate vicinity of the poles themselves.

polarized radiation Electromagnetic radiation in which the electrical field vector is contained in a single plane, instead of having random orientation relative to the propagation vector. Most commonly refers to **radar** images.

principal component analysis The analysis of covariance in a multiple data set so that the data can be projected as additive combinations on to new axes, which express different kinds of correlation among the data.

principal components Directions, or axes, in multispectral space, defined by approximating the data from a whole image, or selected area in an image, to a multi-dimensional ellipsoid. The first principal component direction is parallel to the longest axis of this ellipsoid, and the other principal components are orthogonal to it. The first principal component of the image is the data displayed relative to an axis parallel with the first principal component direction.

principal point The centre of an aerial photograph.

probability density function (PDF) A function indicating the relative frequency with which any measurement may be expected to occur. In remote sensing it is represented by the histogram of digital numbers (DN) in one band for a scene.

pushbroom system An imaging device consisting of a fixed, linear array of many sensors, usually CCDs, which is swept across an area by the motion of the platform, thereby building up an image. It relies on sensors whose response and reading is nearly instantaneously, so that the image swath can be segmented into pixels representing small dimensions on the ground.

quantum (pl. **quanta**) The elementary quantity of electromagnetic (EM) energy that is transmitted by a particular wavelength. According to the quantum theory, EM radiation is emitted, transmitted and absorbed as numbers of quanta, the energy of each quantum being a simple function of the frequency of the radiation.

quantum theory A theory about electromagnetic radiation (EMR) which dictates that there is a minimum amount of energy that can exist (or be emitted) at any particular wavelength. This minimum amount is a **quantum**, or **photon**.

radar The acronym for **radio detection and ranging**, which uses pulses of artificial electromagnetic (EM) radiation in the 1 mm to 1 m range to locate objects which reflect the radiation. The position of the object is a function of the time that a pulse takes to reach it and return to the antenna.

radar altimeter A non-imaging device that records the time of radar returns from vertically beneath a platform to estimate the distance to and hence the elevation of the surface. Carried by Seasat and the ESA ERS-1 platforms.

radar cross-section A measure of the intensity of back-scattered radar energy from a point target. Expressed as the area of a hypothetical surface that scatters radar equally in all directions and which would return the same energy to the antenna.

radar interferometry Use of dual antennae to record backscattered radar energy, so that the antenna separation is analogous to the distance between stereoscopic aerial photographs. Radiation travels different distances from each point on the ground to each antenna, so that when the signals are combined they interfere. Using the geometry of the radar set-up and the known wavelength of the radiation the interference record is converted to topographic elevation of each point. This produces a **digital elevation model** of the overflown surface.

radar scattering coefficient A measure of the back-scattered radar energy from a target with a large area. Expressed as the average radar cross-section per unit area in decibels (dB). It is the fundamental measure of the radar properties of a surface.

radar scatterometer A non-imaging device that records radar energy back-scattered from terrain as a function of depression angle.

radial relief displacement The tendency of vertical objects to appear to lean radially away from the centre of a vertical aerial photograph. Caused by the conical field of view of the camera lens.

radiance The radiant flux density falling on a surface measured per solid angle. A useful concept because we rarely measure the radiant flux density leaving a surface in all directions; instead we collect it with a detector responding to electromagnetic radiation (EMR) arriving at the detector from a finite solid angle.

radiant flux The power incident on or leaving a body in the form of electromagnetic radiation (EMR).

radiant flux density The power incident on or leaving a body in the form of electromagnetic radiation (EMR) per unit area of surface.

ramp stretch See **equalization stretch**.

range In **radar** usage this is the distance in the direction of radar propagation, usually to the side of the platform in an imaging radar system. The **slant range** is the direct distance from the antenna to the object, whereas the distance from the ground track of the platform to the object is termed the **ground range**.

range resolution In a **radar** image, the resolution in the down-range direction, i.e. at right angles to the flight line, as distinct from azimuth resolution. Range resolution depends on the pulse length and the incidence angle, and is not made worse by flying at high altitude.

raster The scanned and illuminated area of a video display, produced by a modulated beam of electrons sweeping the phosphorescent screen line by line from top to bottom at a regular rate of repetition.

raster format A means of representing spatial data in the form of a grid of digital numbers (DN), each line of which can be used to modulate the lines of a video raster.

ratio The digital number (DN) of one band of a multispectral image divided by the DN of another band usually rescaled on to an 8-bit scale. This emphasizes the relative differences between the two channels and ignores brightness/intensity/albedo.

Rayleigh criterion A way of quantifying surface roughness with respect to wavelength of electromagnetic radiation (EMR), to determine whether the surface will act as a specular (smooth) reflector or as a **diffuse** (rough) **reflector**. A surface can be considered rough if the root mean square height of surface irregularities is greater than one eighth of the wavelength of EMR divided by the cosine of the incidence angle.

Rayleigh scattering Selective scattering of light in the atmosphere by particles that are small compared with the wavelength of light.

RBV The Return-Beam Vidicon system aboard **Landsat-3**, which produced panchromatic digital images with a resolution of 40 m.

real-aperture radar An imaging radar system where the azimuth resolution is determined by the physical length of the antenna, the wavelength and the range. Also known as brute-force radar.

receptor Light-sensitive cell embedded in the eye's retina. See **rods** and **cones**.

red edge The sharp increase in spectral reflectance of healthy leafy vegetation, which occurs with increasing wavelength between about 700 nm and 750 nm wavelength, i.e. in the red/very-near-infrared part of the spectrum.

redundancy Information in an image that is either not required for interpretation or cannot be seen. Redundancy may be spatial or spectral. Also refers to multispectral data where the degree of correlation between

bands is so high that one band contains virtually the same information as all the bands.

reflectance/reflectivity The ratio of the electromagnetic (EM) energy reflected by a surface to that which falls upon it. It may be qualified as spectral reflectance. The suffix *-ance* implies a property of a specific surface. The suffix *-ity* implies a property for a given material.

reflected infrared That part of the infrared spectrum that, in terrestrial remote sensing, is mostly made up of reflected solar electromagnetic radiation (EMR), as opposed to the thermal infrared part of the spectrum, which has longer wavelengths. The reflected infrared region is defined as wavelengths from 0.7 μm to 3.0 μm , and is synonymous with the near-infrared.

reflected region That part of the spectrum between 0.4–2.5 μm where most energy is reflected from surfaces.

region of interest (ROI) See *mask*.

resampling The calculation of new digital numbers (DN) for pixels created during geometric correction of a digital scene, based on the values in the local area around the uncorrected pixels.

resolution A poorly defined term relating to the fidelity of an image to the spatial attributes of a scene. It involves the instantaneous field of view (IFOV) and modulation transfer function (MTF) of the imaging system, and depends on the contrast within the image, as well as on other factors. It is usually expressed as line pairs per millimetre of the most closely spaced lines that can be distinguished, and therefore depends on human vision, scale and viewing distance.

RGB colour system Based on the **additive primary colours** red, green and blue. This is the system used by all video monitors in which various data are assigned to red, green and blue colour guns.

ringing Fringe-like **artefacts** produced at edges by some forms of **spatial-frequency filtering**.

rods The **receptors** in the retina that are sensitive to brightness variations.

rotational transition A mechanism of spectral absorption of electromagnetic radiation (EMR), when a photon of the correct wavelength is absorbed to cause a gaseous molecule to undergo a change in its rotational energy state.

saturation In colour theory it means the degree of mixture between a pure hue and neutral grey. Sometimes used to refer to the maximum brightness that can be assigned to a pixel on a display device, and corresponds to a digital number (DN) of 255.

scale In cartography this refers to the degree of reduction from reality that is represented on a map, usually expressed as a ratio (e.g. 1 : 250 000), a representative fraction (e.g. 1/250 000) or an equivalence (e.g. 1 cm = 2.5 km). A large-scale map represents ground dimensions by larger cartographic dimensions than

a smaller-scale map. This is the opposite of the common usage wherein a large-scale feature has larger dimensions than a smaller-scale feature.

scattering An atmospheric effect where electromagnetic (EM) radiation, usually of short visible wavelength, is propagated in all directions by the effects of gas molecules and aerosols. See **Rayleigh**, **Mie** and **non-selective scattering**.

scene The area on the ground recorded by a photograph or other image, including the atmospheric effects on the radiation as it passes from its source, to the ground and back to the sensor.

scotopic vision Vision under conditions of low illumination, when only the rods are sensitive to light. Visual acuity under these conditions is highest in the blue part of the spectrum.

Seasat Polar-orbiting satellite launched in 1978 by NASA to monitor the oceans, using imaging radar and a radar altimeter. It survived for only a few months.

Sea-WIFS Sea Wide-Field Sensor—replacement remote sensing system for the CZCS carried by a satellite launched in 1997.

selective radiator A material that has emissivities that vary with wavelength.

short wavelength infrared (SWIR) That part of the near or reflected infrared region with wavelengths between 1.0 μm and 3.0 μm . These wavelengths are too long to affect infrared photographic film, hence the distinction from VNIR.

Shuttle Radar Topography Mission (SRTM) Interferometric radar mission to map surface elevation of the continents between 60° N and 58° S, with a resolution up to 25 m. DEMs derived from the mission will be publicly available from NASA/JPL and the German Space Agency (DLR) from mid-2000.

signal-to-noise ratio (S/N): The ratio of the level of the signal carrying real information to that carrying spurious information as a result of defects in the system.

signature In remote sensing this refers to the spectral properties of a material or homogeneous area, most usually expressed as the range of digital number (DN) in a number of spectral bands.

SIR Shuttle Imaging Radar, synthetic-aperture radar experiments carried aboard the NASA Space Shuttle in 1981, 1984 and 1994.

slant range Direct distance from a radar antenna to a point at the Earth's surface. Because radar uses energy transmitted obliquely downwards, slant range is different from actual distance on the ground, or **ground range**, and varies through the scene in the look direction.

SMIRR Shuttle Multispectral Infrared Radiometer, a non-imaging spectroradiometer carried by the NASA Space Shuttle covering 10 narrow wavebands in the 0.5–2.4 μm range.

soil line A diagonal line at 45° on a plot of red digital numbers (DN) against very near-infrared (VNIR) DN. Bare soil plots close to this line, but leafy vegetation plots away from this line, having much higher VNIR reflectance.

spatial-frequency filtering The analysis of the spatial variations in digital numbers (DN) of an image and the separation or suppression of selected frequency ranges.

specific heat The ratio of the heat capacity of unit mass of a material to the heat capacity of unit mass of water.

spectral hue A hue that is present in the spectral range of white light analysed by a prism or diffraction grating.

spectral quantities A measure of electromagnetic radiation (EMR) per unit wavelength, as opposed to overall wavelengths. Thus we have spectral radiance, which is radiance per unit wavelength, and so on.

spectral radiant flux The power received or radiated by a body per unit area per unit wavelength. For convenience, this is often quoted in $\text{Wm}^{-2} \mu\text{m}^{-1}$ (**spectral radiance** is the same measure in terms of solid angles as $\text{Wm}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$).

spectral response A general term referring to the amount of light reflected by a surface at different wavelengths, as seen in an image. The spectral response of a surface is controlled principally by its spectral reflectance.

spectroradiometer A device that measures the energy reflected or radiated by materials in narrow electromagnetic (EM) wavebands.

specular reflection Reflection of electromagnetic radiation (EMR) from a smooth surface, which behaves like a mirror. A specular reflector has a surface that is smooth on the scale of the wavelength of EMR concerned. If the surface is rough on the scale of the wavelength of the EMR, it will behave as a **diffuse reflector**.

SPOT Satellite Probatoire pour l'Observation de la Terre, a French satellite carrying two pushbroom imaging systems, one for three wavebands in the **visible** and **very-near-infrared** (VNIR) regions with 20 m resolution, the other producing panchromatic images with 10 m resolution. Each system comprises two devices which are pointable so that off-nadir images are possible, thereby allowing stereoptic viewing. Launched in February 1986 with several later versions.

SRM Shuttle Radar Topography Mission. Carried out in 14 days in February 2000, this Shuttle mission was dedicated to deployment of two **interferometric radar** systems (X- and C-band) aimed at mapping topographic elevation with a 30 m resolution for all land areas between 60° north and south. Data should be available by 2002.

Stefan-Boltzmann Law A radiation law stating that the energy radiated by a **blackbody** is proportional to the fourth power of its **absolute temperature**.

stereopsis The ability for objects to be perceived in three dimensions as a result of the **parallax** differences produced by the eye base.

stereoscope A binocular optical instrument used to view two images with overlapping fields that contain **parallax** differences, as a means of stimulating **stereopsis**.

subtractive primary colours The colours cyan, magenta and yellow, the subtraction of which from white light in different proportions allows all colours to be created.

Sun angle When considering the illumination of a surface by electromagnetic radiation (EMR) from the Sun, the angle between a normal to the surface and the direction of the Sun.

Sun-synchronous orbit A polar orbit where the satellite always crosses the Equator at the same local solar time.

supervised classification A classification technique whereby the human operator identifies training areas on the image that are intended to be representative examples of each surface type.

synthetic-aperture radar (SAR) A radar imaging system in which high resolution in the azimuth direction is achieved by using the **Doppler shift** of back-scattered waves to identify waves from ahead of and behind the platform, thereby simulating a very long antenna.

TDRS Telemetry and Data Relay Satellites, launched by the US Department of Defense into geostationary orbit, primarily for military communications but used for data transfer by Landsat.

Terra The imaginative name chosen through a competition for the first of NASA's **Earth Observing System** platforms, EOS-1AM, after its long-delayed launch in December 1999.

texture The frequency of change and arrangement of tones on an image, often used to describe the aggregate appearance of different parts of the surface, but sometimes used for the spacing of drainage elements.

thematic map An image on which are overlaid the results of classification, usually by showing each class in a distinctive colour. The term **classified image** is synonymous.

Thematic Mapper (TM) An imaging device carried by Landsat-4 to -7, which records scenes in seven wavebands, six in the visible and near-infrared (NIR) regions with a resolution of 30 m, and one in the mid-infrared (MIR) region with a resolution of 120 m. The Advanced Thematic Mapper on Landsat 7 includes a 15 m resolution panchromatic band and improved resolution for its thermal sensor.

thermal capacity The ability of a material to store heat.
thermal conductivity A measure of the rate at which heat passes through a material.

thermal emission Emission of electromagnetic radiation (EMR) from a material owing to thermal vibration as a result of its temperature.

thermal inertia A measure of the response of a material to changes in its temperature. The apparent thermal inertia is calculated from the diurnal change in emitted thermal energy by a material.

thermal infrared (TIR) That part of the spectrum with wavelengths between 3.0 μm and 100 μm . These are the wavelengths at which thermal emission is greatest for surfaces at normal environmental temperatures. However, hot areas (fires, parts of volcanoes and so on) emit thermally at wavelengths shorter than the thermal infrared.

thermal vibration Vibration of atoms or molecules as a result of their temperature. Because these atoms of molecules contain moving electric charges, this results in thermal emission of electromagnetic radiation (EMR).

tie point A point on the ground that is common to two images. Several are used in the co-registration of images.

TIMS Thermal Infrared Multispectral Scanner, an airborne device used by NASA to measure and create images of mid-infrared (MIR) energies emitted by the surface in six wavebands.

TIROS-N The most widely used series of Sun-synchronous meteorological satellites, carrying the AVHRR visible to thermal infrared imaging system.

tone Each distinguishable shade from black to white on a monochrome image, sometimes called grey-tone.

training area A sample of the Earth's surface with known properties, the statistics of the imaged data within which are used to determine decision boundaries in classification.

transmittance/transmissivity The ratio of the electromagnetic (EM) energy passing through a material to that falling on its exposed surface.

transpiration The production and emission of water vapour and oxygen by plants.

Triangular Irregular Network (TIN) A set of irregularly spaced three-dimensional nodes that represent a surface as a set of adjacent, conterminous triangles or facets.

tristimulus colour theory A theory of colour relating all hues to the combined effects of three **additive primary colours** corresponding to the sensitivities of the three types of cone on the retina.

ultraviolet That part of the spectrum with wavelengths between about 0.5 nm and 400 nm. Not useful for most aspects of terrestrial remote sensing, because of scattering in the atmosphere and absorption by ozone.

unsupervised classification A classification technique in which classes are identified by a computer-driven search for clusters in multispectral data space.

variance A measure of the dispersion of the actual values of a variable about its mean. It is the mean of the squares of all the deviations from the mean value of a range of data.

vector format The expression of points, lines and areas on a map by digitized Cartesian co-ordinates, directions and values.

vegetation index A technique, usually involving ratioing, whereby channels from a multispectral image are combined to show up variations in the amount of vegetation. A simple example is to divide **very-near-infrared** digital numbers (DN) by red DN.

Venn diagram Schematic representation of logical relationships between sets of attributes.

vertical exaggeration The extent to which the vertical scale exceeds the horizontal scale in stereoptic viewing of two overlapping images with parallax differences. It is directly proportional to the base height ratio.

very-near-infrared (VNIR) The shortest wavelength part of the near-infrared (reflected infrared), with wavelengths between 0.7 μm and 1.0 μm . Virtually synonymous with **photographic infrared**.

vibrational transitions A mechanism of spectral absorption of electromagnetic radiation (EMR), when the vibration of a chemical bond in a molecule of crystalline lattice changes from one state to another. Vibrations may occur either as stretching or bending of a bond.

vidicon An imaging device based on a sheet of transparent material in which electrical conductivity increases with the intensity of electromagnetic (EM) radiation falling on it. The variation in conductivity across the plate is measured by a sweeping electron beam and converted into a video signal. Now replaced by cameras employing arrays of charge-coupled devices (CCDs).

vignetting A gradual change in overall tone of an image from the centre outwards, caused by the imaging device gathering less radiation from the periphery of its field of view than from the centre. Most usually associated with the radially increasing angle between a lens and the Earth's surface, and the corresponding decrease in the light-gathering capacity of the lens.

visible radiation Electromagnetic radiation (EMR) in that part of the electromagnetic spectrum that human eyes respond to. It lies between the ultraviolet and the infrared regions, with wavelengths from 400 nm to 700 nm.

VISSR Visible Infrared Spin-Scan Radiometer carried by the GOES satellites.

visual dissonance The disturbing effect of seeing a familiar object in an unfamiliar setting or in an unexpected colour.

volume scattering Scattering of electromagnetic (EM) radiation, usually **radar**, in the interior of a material. It may apply to a vegetation canopy or to the subsurface of soil.

wavelength The mean distance between maxima or minima of a periodic pattern. In the case of electromagnetic (EM) radiation, it is the reciprocal of the frequency multiplied by the velocity of light.

whiskbroom Imaging device that sweeps radiant energy from lines on the ground across a CCD array made up of several hundred devices, each designed to record radiance in a narrow spectral band. Used in **imaging spectrometers**.

Wien's Displacement Law A radiation law stating that the peak of energy emitted by a material shifts to shorter wavelengths as absolute temperature increases.

Index

Note: page numbers in *italics* refer to figures; those in **bold** refer to tables.

- absolute temperature, 3
- absorptance, 5, 161
- absorption spectra, 5
- absorptivity, 165
- achromatic axis, 131–2
- active remote sensing, 4
- acuity, 21
 - and colour vision, 21, 23
 - spatial, 23–4
- additive primary colours, 26, 34–5, 219, *Plate 2.1*
- Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), 65, 130, 145
 - spectral bands, 58
 - spectrometers, 66
 - thermal images, 173
- Advanced Very High Resolution Radiometer (AVHRR), 57–8
 - images, 58
 - sensors, 122
 - spectral bands, 58
 - thermal images, 166–7, 169
 - thermal inertia estimation, 171
- aeolian landforms, 114–16
- aerial photographs, 51–2
 - distortions, 36–7
 - elevation data, 214
 - geological applications, 66
 - image interpretation, 36, 68–121
 - infrared, 34
 - minus-blue, 34
 - oblique, 36
 - principal point, 36, 37
 - radial-relief displacement, 36, 37
 - stereoscopic, 37, 68–121, 249, 251
 - vertical, 36, 37
 - vignetting, 37
- aerial photography
 - applications, 50, 51
 - flight lines, 37, 51–2, 249
 - historical background, 50–1
 - overlap, 51
 - sidelap, 51
 - stereoscopic, *Plate 9.1*
- aerial surveys, commissioning, 51
- aeromagnetic data *see* magnetic data
- aesthetics, and colour, 27–8
- Afghanistan, 59
- Africa, 224, *Plate 9.2*
- age, and perceptual ability, 16
- agriculture, monitoring, 61
- airborne data, 51–3
- Airborne Imaging Spectrometer (AIS), 52
- airborne radar, applications, 53
- Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), 52–3, 57, 66, 152, 236
- aircraft
 - advantages, 51
 - cameras, 32
 - disadvantages, 51
 - tilting, 37
- AIS (Airborne Imaging Spectrometer), 52
- Alaska, 217, 220
- albedo, 5, 130, 144, 145, 162, *Plate 6.1*
 - constant, 172
 - estimation, 167
 - and metal exploration, 237, *Plate 5.6*
 - and surface temperature, 163–4
 - and thermal inertia, 165, 166, 167, 169, 171
- albite, 10
- Algeria, 106, 196
- alluvial fans, 112, 114
- alluvium, 113, 120, 241, *Plates 6.2–3, Plate 9.1*
 - radar images, 193
- Al-OH bonds, 9, 142, 152, 172
- altimeters
 - laser, 214
 - radar, 214, 231, 238
- aluminium, energy level, 7
- American Civil War (1861–5), 50
- amphibolites, *Plate 8.4*
- anaglyphs, 250
- Anaimalai Hills (India), 103
- analogue-digital conversion, 122
- Andes, 124, 243, 257
 - volcanoes, 103, 108, 109, 111, 112
- Andhra Pradesh (India), *Plate 4.2*
- AND operator, 222
- annular drainage, 72, 73
- anorthite, 10
- antecedent drainage, 70, *Plate 4.1*
- antennae
 - radar, 43, 46
 - Space Shuttle, 204
- anticlines, 104, 170
- apatite, 11
- Apollo missions, 50, 55
- Appalachian Plateau (US), 55, 59, 181, 199
- apparent thermal inertia (ATI)
 - estimation, 171
 - images, 171
- Applications Technology Satellite (ATS-1), 57
- Arctic (Canada), 113
- Argentina, 111, 112
- Arizona (US), 70, 86, 108, 115, *Plate 4.1*
- artefacts, 96, 138
- artificial intelligence, and pattern recognition, 145, 147
- arêtes, 74
- ash cones, 110
- ASTER *see* Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)
- ATI *see* apparent thermal inertia (ATI)
- atmospheric correction, 125–7
- atmospheric effects, 6–7, 140, 228
 - gas absorption, 6
 - haze, 126
 - and ratioing, 143–4
 - scattering, 6–7, 17, 127, 128
 - shimmer, 7
- atmospheric windows, 7, 9
- atoms, coordination states, 4
- ATS-1 (Applications Technology Satellite), 57
- augite, 10
- Australia, 89, 182, *Plate 8.4, Plate 8.6*
- AVHRR *see* Advanced Very High Resolution Radiometer (AVHRR)
- AVIRIS (Airborne Visible/Infrared Imaging Spectrometer), 52–3, 57, 66, 152
- backscatter (radar)
 - blooming, 180–6
 - coefficient, 176, 178
 - dielectric constant, 176–7
 - mechanisms, 178
 - and polarization, 180
 - and roughness, 177–9, 187, 189
 - surface materials, 176
- bacteria, anaerobic, 239
- Badain Jaran (Mongolia), 190
- balloons, aerial photography, 50
- band-pass filters, 134
- band ratioing *see* ratioing
- band-stop filters, 134
- barchans, 114–15
- baryte, 222
- basalts, *Plate 6.4*
 - flood, 85, *Plate 9.1*
 - outcrops, 222
 - plateau, 85
 - raster images, 219
 - reflectance, 147
- basins, 107
- batholiths, 93
- bathymetry, 214, 231, *Plate 1.1*
- bedding, 76–7
- beryl, 8
- best-fit lines, 209
- binary numbers
 - in data transmission, 32
 - in image processing, 122
- bismuth, anomalies, 237
- bivariate plots, 139, 143, 150, 152, 153
- blackbodies, 138, 169
 - emittance, 3–4, 10, 160
 - water as, 13–14
- blind spot, 17
- blooms (radar), 180–6
- blue sky, 6–7
- bond bending, 9
- bonds, distortions, 4
- Boolean algebra, 222
- Borrowdale (UK), *Plate 9.6*
- Bouguer gravity anomalies, 218, 221, *Plate 8.5, Plate 9.6*
- box filters, 134
- Brahmaputra, 70
- braided streams, 113, 119
- brightness, 2
 - and contrast, 25, 27
 - detection, 23, 24
- bronzite, 8
- buried landscapes, 95
- buttes, 77
- byte-format data, 122

- calcite, 10, 11
 calderas, 102, 110–12, 188
 Caledonides (Scotland), 99
 California (US), 101, 194, *Plate 5.7*, *Plate 9.9*
 Mojave Desert, 167, 172, 193, *Plate 6.3*
 camera lenses
 aberration, 33
 field of view, 36
 refraction, 33
 cameras
 on aircraft, 32
 aperture, 17, 33
 eye analogy, 16–17
 film speed, 33
 focal length, 17, 33
 f stop, 17, 33
 on manned spacecraft, 55
 metric, 36, 56
 video, 32
see also digital cameras; Vidicon cameras
 Camerouns, 86
 camouflage, 242
 detection, 147
 reflectivity, 34
 Canada, 118, 119, 120, *Plate 9.3*
 Canadian Arctic, 113
 Canadian Shield, 75, 91, 98
 cancers, 243
 canonical analysis, 141, 152
 capital investment, and exploration, 234
 carbonate minerals, 83, *Plates 6.4–5*
 radar images, 183
 spectra, 9, 10, 130, 142
 thermal inertia, 165
 weathering, 82
 carbon dioxide, 6, 160
 cartography
 digital, 66–7, 233
see also maps
 catchments, 211, 212, 223, 240
 CCDs *see* charge-coupled devices (CCDs)
 CD-ROM resources, 259–63
 CD-ROMs, data storage, 66
 cements, 77, 79
 Centre National d'Études Spatiales (CNES), 62
 centripetal drainage, 72, 73
 change detection, 156–7
 automated, 157
 charge-coupled devices (CCDs), 38, 42, 62, 63
 in cameras, 17
 in imaging spectrometers, 49
 charge-transfer transitions, 8
 Chile, 108, 109, 111, 257
 China, 183, 189, *Plate 4.4*
 chlorite, 9
 chlorophyll, 12–13, 61, 144, 145
 chromaticity diagrams, 28, 29
 chromium
 anomalies, 219
 concentration, 223, *Plate 8.3*
 CIE *see* Commission International de l'Éclairage (CIE)
 cinder cones, 108, 109, 110
 circular features, detection, 102–4
 cirques, 74
 classification
 geological applications, 151–2
 methods, 149–51, 153–4, 156
 multispectral, 147–52
 problems, 151
 and radar data, 199–200
 surface materials, 147–52, 169
 thermal data in, 171
 clastic sediments, 192
 coarse, 77–8
 erosion, 79
 fine, 77, 78–81
 radar images, 183
 clay minerals, 78–9, 80–1, *Plate 6.4*
 detection, *Plate 3.2*
 discrimination, 151
 isotopes, 219
 and ratio images, 145
 reflectance, 141
 spectra, 9, 236
 thermal inertia, 165
 clayrocks *see* mudstones
 cleavage, metamorphic rocks, 90
 climate
 and geomorphology, 229
 and water supplies, 240
 CNES (Centre National d'Études Spatiales), 62
 coalfields, *Plate 9.9*
 Coastal Zone Color Scanner (CZCS), 58
 coasts, geological interpretation, 75
 C–O bonds, 9, 10, 142, 152, 172
 coherent radiation, 43
 Colorado Plateau (US), *Plate 4.1*
 Colorado River (US), 70, *Plate 4.1*
 Colorado (US), 165, 168
 colour atlas, 28
 colour cubes, 131, 132
 colour images
 band choice, 129–31
 colour range, 122
 contrast stretching, 131–3
see also red–green–blue (RGB) images
 colour perception, 145
 colour photography, 34–6, 68
 colour representation, in infrared images, 27
 colour theory
 Munsell's, 28
 tristimulus, 28
 Young's, 26–7, 218
 colour transformations, 219–20
 colour vision, 17–18, 26–8
 and acuity, 21, 23
 and aesthetics, 27–8
 early studies, 26
 image memory, 27
 physiological effects, 27
 psychological effects, 27
 quadrichromatic, 26
 trichromatic, 26
 colour wheels, 27, 132, *Plate 5.2*
 Commission International de l'Éclairage (CIE)
 chromaticity diagram, 28, 29
 colour co-ordinate system, 28, 133
 tristimulus values, 28
 cones (vision), 17–18, 21, 26–7
 tonal sensitivity, 26
 types of, 18, 26
 coniferous trees, 13
 consequent streams, 70, 93
 conservation of energy, 5
 constructional landforms, 69, 107–20
 context, 69
 contextual classification, 156
 continental glaciation, 74–5, 116, 117, 119
 contour maps, 30, 215
 digitizing, 211
 elevation data, 213–14
 sampling, 210–11
 visual perception, 206
 contours, optical illusions, 30
 contrast, 21, 24
 and brightness, 25, 27
 contrast stretching, 123, 125–33, 137
 colour images, band choice, 129–31
 in colour spaces, 131–3
 decorrelation, 131, 138, 141, 142–3
 equalization, 128
 exponential, 129
 Gaussian, 128
 goals, 125
 linear, 125
 logarithmic, 128
 magnetic data, 215
 nonlinear, 128
 piecewise, 128
 ramp, 128
 in thermal images, 168
 convolution, 134–5
 limitations, 137–8
 matrices, 135, 136, 137
 copper
 anomalies, *Plate 9.6*
 concentration, 212, 223
 mineralization, 222–3, 237
 coral reefs, *Plate 1.1*
 corn, spectra, 146
 cornea, 16
 corner reflectors, 177, 180–6, 227
 Corona satellites, 51
 correlation coefficients, 139
 cortex (visual) *see* visual cortex
 covariance, 139
 covariance matrices, 139
 cover–basement relationships, 217
 craters, 102–4, 232
 optical illusions, 30
 cross-cutting relationships, 95
 cruise missiles, guidance systems, 214
 crustal deformation, 96–7, 104–7
 crystal-field effects, 8
 cubic splines, 209, 210
 Cuddapah Basin (India), 95, *Plate 4.2*
 cuestas, 91–2, 93
 Cumberland Plateau (US), 91
 cut-off (image correction), 125
 CZCS (Coastal Zone Color Scanner), 58
 daguerreotype photography, 50
 dams, and water supplies, 240–1
 Danakil Depression (Ethiopia), 201
 dark-pixel corrections, 125
 Dasht-e-Lut (Iran), 169
 data collection, images, 32–67
 data reduction, 131, 138–45
 data sources, 264–5
 data storage
 CD-ROMs, 66
 elevation data, 213
 magnetic tape, 39
 data transmission
 analogue, 32
 digital, 32
 Davis, W. M., 70, 228, 229
 deciduous trees, 13
 decorrelation stretching, 131, 138, 141, 142–3
 Delaunay triangles, 213

- deltas, 114, 119
- DEMs *see* digital elevation models (DEMs)
- dendritic drainage, 72, 73, 86, 88, 91, 94
 - development, 79–80
- density slicing, 149
- depression angles, 43, 66
 - and radar images, 176, 178, 180, 195–7
- depth perception, 29–30
- deserts, radar images, 187, 190, 192, 193, 196
- desert varnish, 148, 187
- de-striping, 255–7
- destructural landforms, 69–76
- detectors
 - area-array, 52
 - faults, 42
 - photoelectronic, 39
 - photosensitive, 38
 - radiation-sensitive, 38
 - sensitivity, 39
 - signal-to-noise ratio, 39
- Devon (UK), *Plate* 3.1
- dew, 162
- diapirs, 88
- dielectric constant, surface materials, 176–7, *Plate* 7.3
- differentiation, in raster data processing, 217
- diffraction gratings, 38
- diffuse reflection, 5–6
- digital cameras
 - charge-coupled devices, 17
 - data, 66–7
- digital cartography, 66–7, 233
- digital elevation models (DEMs), 66, 204, 213–14, 216, 252
 - engineering applications, 241
 - forms, 213
 - geomorphological applications, 223–4, 229
 - hydrological applications, 223–4
 - and parallax, 220, 233
 - and perspective views, *Plate* 9.2
- digital image processing, 122–59
 - goals, 122
- digital images
 - applications, 66–7
 - data reduction, 131
 - directional bias, 137
 - high-frequency features, 133
 - histograms, 123–5
 - low-frequency features, 133
 - redundancy, 131
 - registration, 122
- digital numbers (DNs), 32, 39–40, 122
 - in colour displays, 218
 - in digital image processing, 123, 125–8, 129
 - in image correction, 254
 - in noise removal, 257
 - in ratioing, 143
 - in spatial-frequency filtering, 133–6
 - spatial frequency variations, 155, 156
 - in thermal images, 160, 169
- digital scanning, film, 32
- Digital Terrain Elevation Data (DTED), 213
- dip, 91–3, 213, 233, 250–1
 - faults, 105–6
- directed principal component analysis, 141–2
- directional bias, 137
- directional filtering, 136–7, 138
- distance estimation, 29
- diurnal heating cycle, 163–4
- DLR (German Space Agency), 204
- DNs *see* digital numbers (DNs)
- dolomite, 10, 83
 - thermal inertia, 165, 166, 171
- domes, 107
- Doppler shift, 46, 176
- drainage patterns, 70–3, 224, *Plate* 4.1
 - categorization, 230
 - tributaries, 223
 - see also* dendritic drainage; trellis drainage
- drainage surveys, 211–12, 244
- dropped lines, replacing, 255
- drumlins, 117, 118
- dryness, and radar images, 188–9
- D-stretching *see* decorrelation stretching
- DTED (Digital Terrain Elevation Data), 213
- ductile faults *see* shear zones
- dunes, 114–16, 190
- dykes, 95, 96, 97, 138
 - basaltic, *Plate* 5.1, *Plate* 5.9
 - detection, 87, 240–1, *Plate* 4.2
 - joints, 85
- Earth
 - ellipsoid, 253
 - geoid, 253
 - internal heat flow, 11
 - microwave emission, 176
 - radiance, 161
 - surface geometry, 253–4
 - surface heat removal, 161–2
 - surface temperature, 4, 11, 43, 160
 - profiles, 163–4
- Earth Observation Satellite Corporation (EOSAT), 59
- Earth Observation System (EOS), 142
 - future prospects, 66
 - launch, 65
 - platforms, 66
 - thermal images, 173
- earthquakes, 99–100, 242, *Plate* 9.9
- Earth Resources Technology Satellite (ERTS-1) *see* Landsat
- Earthwatch, 66
- East Africa, 216
- East African Rift (Kenya), *Plate* 9.8
- edges
 - detection, 135
 - enhancement, 135–6
 - representations, 133–4
- Egypt, 193
- eigenvalues, 139–41, 153
- eigenvectors, 139–41, 142, 153
- Einstein's equation, 1
- electromagnetic (EM) spectrum, 2, 3, 6
- electromagnetic radiation (EMR), 1–15
 - absorption, 5
 - atmospheric effects, 6–7
 - characteristics, 1–2
 - emitted region, 7
 - fields, 1
 - generation, 2–4
 - intensity, 2
 - and matter, 4–14
 - microwave region, 7, 11, 43, 176
 - polarized, 1–2
 - reflected region, 7
 - reflection, 5
 - reradiation, 5
 - rock interactions, 7–11
 - scattering, 5, 6–7
 - transmission, 5
- vegetation interactions, 12–13
- water interactions, 13–14
- electronic transitions, 4
- elements
 - concentrations, 211–12
 - energy levels, 7–8
- elevation data, 213–14, 233, 252
- emission spectra, 5
- emissivity, 10, 160, 165
- emittance, 2, 3–4, 161
- EMR *see* electromagnetic radiation (EMR)
- EM spectrum *see* electromagnetic (EM) spectrum
- energy, 2
 - conservation of, 5
 - detection, 4–5
- engineering applications, 241–3
- England, 82, 215, 218, *Plate* 8.1, *Plate* 8.5
- Devon, *Plate* 3.1
- Lake District, 237, *Plate* 9.6
- Selby Coalfield, *Plate* 9.9
- environmental hazards
 - geochemical, 243–4
 - geophysical, 241–3
- Environmental Science Service (ESSA), satellites, 57
- EOS *see* Earth Observation System (EOS)
- EOSAT (Earth Observation Satellite Corporation), 59
- equalization stretch, 128
- Eritrea, 56, 99, *Plate* 3.3, *Plate* 9.1, *Plate* 9.5
- Mersa Fatma, *Plate* 1.1
- western lowlands, 102
- see also* Sudan–Eritrea border
- Eritrean Highlands, 256
- EROS Data Centre, 55
- erosion, 69–70
 - cycles, 70
 - glacial, 73–5, 83
 - headward, 70, 72
 - marine, 69, 75
 - rates, 72
 - soils, 240
 - stream, 70–2
 - wind, 69, 75–6
- erosion resistance, 69, 76–7, 90
- ERS-1 (satellite), 64, 230–1
 - images, 65
 - radar altimeters, 214
- Erta' Ale volcano (Ethiopia), 201, *Plate* 7.2
- ESA *see* European Space Agency (ESA)
- escarpments, 91–2, 93
- eskers, 117
- ESSA (Environmental Science Service), satellites, 57
- ethics, and exploration, 244
- Ethiopia, 201, *Plate* 7.2
- Ethiopian Highlands, 255
- Eureka Valley (US), *Plate* 9.9
- European Space Agency (ESA), 57
 - satellites, 64, 204
- evaporites, 81–3
- evapotranspiration rates, 240
- exfoliation, 88
- exitance, 2
- exploration, 233–41
 - and capital investment, 234
 - economics, 234, 235
 - ethical issues, 244
 - goals, 234
 - market-led, 234
 - offshore, 239
 - risks, 234

- exploration (*cont.*)
 - stages, 234, 235
 - targeting, 234–6
 - see also* geochemical exploration; hydrocarbon exploration; metal exploration; water exploration
- exponential stretch, 129
- extrapolation, 208
- eye base, 29
- Eyeglass, 66
- eyes
 - compound, 16
 - evolution, 16
 - see also* human eye
- faceted spurs, 100, 101
- false-colour images, 35–6, 126–7, 228, 230
 - band choice, 129–31, *Plate 3.1*
 - thermal data in, 168–9
- fan drainage, 72
- faults, 95–104, *Plate 8.4*, *Plate 9.1*
 - active, 100
 - angles, 97
 - and construction projects, 242
 - detection, 97, 98, 240
 - issues, 224
 - as environmental hazards, 99–100, 242
 - and geographical information systems, 224
 - mapping, 233
 - normal, 96, 97, 98
 - radar images, 199
 - reverse, 96, 97
 - splay, 97
 - strike-slip, 96, 97, 98, 99, 101, 102
 - thrust, 96, 97
 - types of, 96
- FCCs *see* false-colour images
- feldspars, 84, 86, 236
- Fe–O bonds, 172
- ferric minerals *see* iron minerals
- ferromagnesian minerals, 84, 236
- field studies
 - economics, 235
 - and remote sensing, 227
- film
 - black and white, 33–4
 - colour
 - infrared, 35
 - natural, 35
 - sensitivity, 35
 - structure, 35
 - digital scanning, 32
 - exposure, 33
 - holographic, 55
 - infrared, 34
 - panchromatic, 34, 56
 - resolution, 33
 - sensitivity, 34
 - speed, 33
 - storage, 32
 - problems, 66
 - structure, 33
- filters
 - band-pass, 134
 - band-stop, 134
 - box, 134
 - coloured, 26, 34
 - directional, 136–7, 138
 - frequency-domain, 134
 - high-pass, 134, 135, 136
 - image-processing, 23, 37
 - low-pass, 134, 135
 - median, 143, 257
 - medium-pass, 134
 - spatial, 134–8
 - square, 138
 - see also* spatial-frequency filtering
- fire fountains, 110
- first derivative filters, 136
- First World War, aerial photography, 50, 51
- flat-field calibration, 154
- flatirons, 47
- flight lines, 37, 51–2, 249
- spectroradiometry, *Plate 3.2*
- flood basalts, 85, *Plate 9.1*
- flood plains, 72, 113–14
- fluorescence, 4, 12
- fluorite, 222
- fluorosis, 243
- fluvial landforms, 112–14
- folding
 - metamorphic rocks, 90, *Plate 4.5*
 - radar images, 199
 - unconformities, 95
- folds, 104–7
 - classification, 104
 - dip, 105–6
 - nomenclature, 105
- foreshortening, 48, 49
- Fourier analysis, 23, 134, 209
- Fourier synthesis, 23
- Fourier transforms, 134
- fovea, 17, 18, 24, 29
- fracture analysis, 224
- fractures
 - and geographical information systems, 224
 - and water supplies, 241
- frequency, 1
- frequency-domain filtering, 134
- Front Range (US), 165, 168
- fundamental wavelength, 4
- Fuyo-1 *see* Japanese Earth Resources Satellite (JERS-1)
- gabbros, 90, 98, *Plate 5.9*
- GAC (Global Area Coverage), 58
- gamma-rays, 4
 - detection, 6
 - generation, 2
- gamma-ray spectrometers, 49–50, 219
- gases
 - radiation absorption, 6
 - volcanic, 110
- Gaussian stretch, 128
- Gemini programme, 55
- geobotany
 - anomalies, 156, 237
 - indicators, 145
 - and ratioing, 145
- geochemical data, 211–13, *Plate 8.3*
 - anomalies, 212, 233–4
 - background threshold, 212
 - drainage surveys, 211–12
 - histograms, 212–13
 - ratioing, 219
 - risk assessment, 243–4
 - sampling, 211–12
- geochemical exploration, 211–12, 220
 - pathfinder elements, 219
- geochemical hazards, 243–4
- geochemical stress, 145
- geochemical surveys, soils, 243–4
- geodetic datum, 253–4
- geographical information systems (GIS), 206–26
 - data analysis, 221–4
 - and digital elevation model, 223–4
 - in exploration, 234–6
 - fault studies, 224
 - functions, 221
 - geological applications, 221–2
 - goals, 221
 - layered structures, 206
 - measurement functions, 221–2
 - and metal exploration, 237
 - multivariate data, 206–7
 - neighbourhood functions, 223
 - organizational issues, 221
 - overlay functions, 222–3
 - raster-based, 207–8, 221, 225
 - registration in, 221
 - retrieval functions, 221–2
 - vector-based, 207–8, 221, 225
- geoid, 253
- geological applications, 50, 51, 56, 58, 227–47
 - aerial photographs, 66
 - classification, 151–2
 - data processing, 228
 - digital images, 66–7
 - geographical information systems, 221–2
 - potential, 64
 - remote sensing strategies, 227–8
 - spatial pattern recognition, 155–6
 - and spectral coverage, 228
- geological mapping, 231–3, *Plate 8.3*, *Plate 8.4*, *Plate 9.3*, *Plate 9.6*
 - boundaries, 231, 233
 - image units, 231–2
 - lithofacies units, 231, 232–3
 - subsurface, 206
- geometric correction, 253–5
- geomorphological applications, 228–31
 - change detection, 157
 - and digital elevation models, 223–4, 229
 - terrain analysis, 228, 230
- geomorphological images, interpretation, 230
- geomorphological mapping, 229–31
 - multitiered approach, 230
- geomorphological theories
 - bottom upwards approach, 228–9
 - climatic, 229
 - Davis's, 228, 229
 - megascopic, 228
 - stageist, 229
 - structural, 229
 - top downwards approach, 229
- geomorphological units, mapping, 230
- geomorphology, 228–31
 - coastal, 75
- geophysical data, 210–11
 - gridding, 210
 - line surveys, 210
 - sampling strategies, 210–11
- Geoscan system, 52, 173
- geostationary satellites, 53–4
 - resolution, 54
- geosynchronous orbits, 53–4
- geothermal heat, 160–1
- German Space Agency (DLR), 204
- GIS *see* geographical information systems (GIS)
- glacial erosion, 73–5, 83
- glacial landforms, 116–20
 - erosional, 73–5
 - radar images, 187

- glaciation
 continental, 74–5, 116, 117, 119
 valley, 73–4
 Global Area Coverage (GAC), 58
 global warming, 6, 66
 gneisses, *Plate 8.3*
 Goddard Space Flight Center (US), 55, 57
 GOES (satellite), 57
 goethite, 8, 145
 goitre, 243
 gold mineralization, 219, 220
 GOMS (satellite), 57
 gossans, 236
 gradient change, 136
 Grand Canyon (US), 70, *Plate 4.1*
 granites, 85–8, 89, *Plate 8.4*, *Plate 9.2–3*, *Plate 9.4*
 anomalies, 237
 health risks, 243
 intrusions, 87, 102
 isotopes, 219
 radar images, 185
 raster images, 219
 reflectance, 147
 and sandstones compared, 86–8
 thermal inertia, 165, 166
 weathered, 148
 gravitational potential, 217
 gravity data
 anomalies, 218, 219, 221, *Plate 8.2*, *Plate 8.5*, *Plate 9.6*
 from ocean-surface elevation, 214
 pseudocolours, 216
 greenhouse effect, 6, 160
 greybodies, emittance, 10
 grey scales, 24, 25, 28, 123
 human discrimination, 125, 215, 228
 gridding, 208–10
 contour maps, 210–11
 geophysical data, 210
 kriging, 210
 moving averages, 210
 regression analysis, 209
 splines, 209–10
 trend-surface analysis, 209
 ground control points, 254
 ground stations, 62, 64
 groundwater
 abstraction, *Plate 9.9*
 accumulation, 240–1
 and subsidence, 242–3
 gullies, 72, 77, 80, 81, 116
 Guyana, 198
 gypsum, 9, 11
 Hammersley Range (Australia), 182
 hanging valleys, 74
 harmonics, 4
 hartzburgite, *Plate 5.9*
 Haute Resolution Visible (HRV), 62, 63
 Hawaii, 109
 hazard assessment *see* risk assessment
 haze, 126
 HCMM *see* Heat Capacity Mapping Mission (HCMM)
 headless valleys, 100
 headward erosion, 70, 72
 health blackspots, 243
 heart disease, 243
 Heat Capacity Mapping Mission (HCMM), 64
 thermal images, 168–9
 thermal inertia estimation, 167, 169–71
 heat transfer, radiative, 161
 Hei-ch'a Shan (China), 189
 hematite, 8, 11, 145
 Herculanum (Italy), 243
 High Atlas Mountains (Morocco), *Plate 6.2*
 Highlands (Scotland), 74
 high-pass filters, 134, 135, 136
 High-Resolution Imaging Spectrometer (HIRIS), 53, 66, 131, 142
 High Resolution Picture Transmission (HRPT), 58
 Himalayas, 59
 Himawari (satellite), 57
 HIRIS (High-Resolution Imaging Spectrometer), 53, 66, 131, 142
 HIS system *see* hue–intensity–saturation (HIS) system
 histograms, 123–5
 in classification, 147, 148, 149, 151
 cumulative, 123
 H–O–H bonds, 9
 hornblende, 10
 HRPT (High Resolution Picture Transmission), 58
 HRV (Haute Resolution Visible), 62, 63
 Huang He (China), 189
 hue, 132
 hue–intensity–saturation (HIS) system, 28, 138, 143, 168, *Plates 5.2–3*, *Plate 7.3*
 transforms, 132–3, 144, 233, *Plate 5.6*, *Plates 8.5–6*, *Plates 9.3–4*
 non-image data, 216, 219–20
 human eye, 16–19
 angular resolution, 30
 blind spot, 17
 camera analogy, 16–17
 chromatic aberration, 21
 cornea, 16
 cross-section, 16
 depth of field, 17
 grey-scale discrimination, 125, 215, 228
 lens, 16
 focal length, 17
 modulation transfer function, 22–3
 parallax, 249
 responses, 21–2
 sensitivity, 18
 spatial acuity, 23–4
 spherical aberration, 21
 see also pupil; retina
 human vision, 16–31
 acuity, 21
 binocular, 29
 brightness detection, 23, 24
 colour perception, 145
 contour map perceptions, 206
 depth perception, 29–30
 and image dissection, 19–21
 and memory, 19
 night, 18
 photopic, 18, 21
 scotopic, 21
 spatial frequency response, 134
 spatial resolving power, 19–24
 stereoptic, 29
 stereoscopic, 19, 249
 see also colour vision; optical illusions
 humidity, 162
 Hyderabad (India), 64
 hydrocarbon exploration, 234, 238–9
 geochemical data, 238–9
 geophysical data, 238
 magnetic anomalies, 239
 offshore, 239
 radar altimetry, 238
 hydrocarbon resources
 formation, 238
 occurrence, 238
 and pollution, 239
 seepages, 238–9, *Plate 9.7*
 hydrogen ions, 236
 hydrogen sulphide, 239
 hydrological mapping, and digital elevation models, 223–4
 hydrothermal fluids, 84, 101, 102, 237
 hydrothermal mineralization, 222
 hydroxides
 and ratio images, 145
 spectra, 8, 130, 142
 hydroxylates, spectra, 9
 hydroxyl-bearing minerals, spectra, 156, 236, *Plate 5.6*, *Plate 5.8*, *Plate 9.5*
 hydroxyl (OH⁺) ions, vibrational transitions, 9
 Hyperion system, 66
 hyperspectral data, 66, 145, 147, 152–5
 ice
 radar images, 186, 187
 spectra, 14
 striations, 74
 ice dams, 119
 ice sheets, 73, 74–5, 118–19
 Idaho, 110
 IFOV *see* instantaneous field of view (IFOV)
 igneous rocks
 colour, 84
 extrusive, 84–5
 intrusive, 85–8
 radar images, 193
 recognition, 83–8
 silicon dioxide content, 174
 thermal inertia, 171, 173
 thermal spectra, 173, 174
 ignimbrites, 84, 110, 111, 112
 image atlases, 264
 image correction, 253–8
 defects, 253
 distortions, 253–5
 image data sources, 264–5
 image dissection
 and human vision, 19–21
 and perceptual quality, 20
 image interpretation
 aerial photographs, 36, 68–121
 geomorphological images, 230
 and optical illusions, 30–1, 96
 precautions, 96
 psychophysiology, 27
 and spatial resolution, 19–24
 surface characteristics, 68
 images
 data collection, 32–67
 data sources, 264–5
 geological applications, 227–47
 geomorphological, 230
 rectification, 254
 redundancy in, 19–20
 resampling, 254
 seasonal, 69, 156, *Plate 5.10*
 sonar, 231
 see also colour images; digital images; false-colour images; infrared images; radar images; red–green–blue (RGB)

- images (*cont.*)
 - images; stereoscopic images; thermal images
- image segmentation, 156
- image striping, 255
- imaging spectrometers, 43, 49, 52–3
 - specifications, 53
 - spectral range, 52
- impact structures, 102–4
- India, 91, 105, 144, 187, *Plate 4.2*
- Indiana (US), 117
- Indian Remote Sensing Satellite (IRS-1), 64
- Indus, 70
- infiltration, 240
- infrared images, 160–74
 - colour representation, 27, 50
 - see also* thermal images
- infrared photography, 34
 - colour, 50
- infrared radiation, generation, 2
- InSAR (interferometric synthetic-aperture radar), 203, 204, 242, 243
- INSAT (satellite), 57
- inselbergs, 88
- insequent streams, 70
- instantaneous field of view (IFOV), 39, 40, 62
 - in geostationary satellites, 54
 - in microwave imaging, 43
- intensity, 2, 132
- intensity–saturation–hue (ISH) system *see* hue–intensity–saturation (HIS) system
- interferometric synthetic-aperture radar (InSAR), 203, 204, 242, 243
- interferometry, *Plate 9.9*
 - radar, 203–4, 213–14, 242
- Internet, data sources, 264–5
- interpolation, 208–10
- intrusions, 85–8, 93, 95
 - granites, 87, 102
- Io, remote sensing, 6
- ionic bonding, 7–8
- Iran, 101, 103, 169, 188
- Irian Jaya, 184, 186
- iron minerals
 - colloidal, 236–7
 - detection, 236
 - discrimination, 151
 - and ratio images, 145
 - reflectance, 141
 - spectra, 8, *Plate 5.6*, *Plate 9.5*
- iron oxides, 148
 - spectra, 8, 130, 142
- iron sulphides, 81
- irradiance, 2
- IRS-1 (Indian Remote Sensing Satellite), 64
- ISH system *see* hue–intensity–saturation (HIS) system
- isotopes
 - abundance, 219
 - gamma-rays, 49–50
 - unstable, 49
- Italy, 243, *Plate 9.9*
- Japan, 48
- Japanese Earth Resources Satellite (JERS-1), 145
 - applications, 230–1, 236
 - data, 62–4
 - false-colour limitations, 129
- images, 65
 - sensors, 62–4
 - spectral bands, 58
 - stereoscopic images, 214
- jarosite, 8
- Jebel al Druz (Syria), 191
- JERS-1 *see* Japanese Earth Resources Satellite (JERS-1)
- joints, 96
 - carbonate minerals, 83
 - columnar, 85
 - dykes, 85
 - granites, 86
 - intrusions, 85
 - lavas, 84
 - sandstones, 77, 79
- Jordan, 137
- Jupiter, moons, 6
- jökulhlaups, 110
- Kaibab Plateau (US), 70, 115
- Kalpin Tagh mountains (China), 183
- kames, 147, 18
- Kansas (US), 78
- kaolinite, 9, 11, *Plate 9.2*
- Karakoram Range (Asia), 187
- Karhunen-Loeve transform *see* principal component analysis (PCA)
- Kentucky (US), 55, 59, 82, 91
- Kenya, *Plate 9.8*
- kettle holes, 116, 119
- Keyhole satellites, 51
- kimberlite pipes, 102
- Kirchhoff's law, 10, 161
- Kirthar Range (Pakistan), 101
- Kiruna (Sweden), 62
- Krige, D. G., 210
- kriging, 217
 - in gridding, 210
- Kuhha-ye Sahand (Iran), 188
- Labrador (Canada), 103
- labradorite, 10
- La Brea (US), 238
- LAC (Local Area Coverage), 58
- lacustrine landforms, 112–14
- lahars, 243
- Lake District (UK), 237, *Plate 9.6*
- lakes, *Plate 7.2*
 - drainage, 114
 - glacial, 119
 - oxbow, 113, 114
- Lambertian reflectors, 5
- landforms
 - aeolian, 114–16
 - constructional, 69, 107–20
 - destructional, 69–76
 - division, 230
 - and drainage, 70–3
 - fluvial, 112–14
 - geomorphological mapping, 230
 - lacustrine, 112–14
 - in oceans, 231
 - see also* glacial landforms
- Landsat, 59–62, 220–1, 229
 - applications, 61
 - development, 55, 59–60
 - future modifications, 62
 - future prospects, 66
 - ground stations, 62
 - Multispectral Scanner, 61, 68, 129, 139, 141, 168–9
 - orbits, 60
- Return-beam Vidicon, 61, 68
- seasonal studies, 156
- stereoscopic potential, 62
- Thematic Mapper, 9, 61–2, 138, 142, 145, 153, 166, 237
- thermal data, 61
- landscape change, 229
- landslides, risk assessment, 241
- Large-Format Camera (LFC), 56, 229
- laser altimeters, 214
- lasers, in remote sensing, 12
- Las Vegas (US), *Plate 9.9*
- lava flows, 107–12
- lavas, 84–5, 124, 188, 191, *Plate 7.2*
 - dissected, 86
 - environmental hazards, 243
 - radar images, 188
 - thermal inertia, 167
 - viscosity, 85
 - weathered, 86
- law of V's, 93, 105, 233
- layovers, 47, 48
- leaves
 - dielectric constant, 177
 - senescence, 145, *Plate 9.7*
 - spectra, 12, 13, 146
- Lewis satellite, 66, 142
- LFC (Large-Format Camera), 56, 229
- Libya, 195
- light, detection, 24
- light scattering, in water, 13
- limestones, 81–3, 84, 98, *Plate 5.9*
 - drainage patterns, 83
 - and health risks, 243
 - hydrothermal mineralization, 222
- isotopes, 219
- radar images, 200
- thermal inertia, 165, 166, 171
- limonite, 8, 81, 83
- lineaments, 95–104
 - use of term, 96
- linear, use of term, 96
- linear features, precautions, 96
- Linear Imaging Self Scanning (LISS) cameras, 64
- linear stretch, 125
- linear unmixing, 154
- line defects, drop outs, 253
- lines, optical illusions, 31
- line scanners, 38–42
 - aerial, 52
 - analogue data, 39
 - arrangement, 38
 - detectors, 38, 42
 - sensitivity, 39
 - digital data, 39–41
 - images
 - analogue, 41
 - defects, 42
 - digital, 41
 - distortions, 41
 - structure, 40–1
 - limitations, 42
 - resolution, 40–1
 - radiometric, 39
 - spatial, 39, 42
 - spectral, 39, 42
 - signal-to-noise ratio, 39
 - telemetry, 39
 - in thermal inertia estimation, 169–71
- line surveys, 210

- LISS cameras (Linear Imaging Self Scanning cameras), 64
- lithic surfaces, 97
- lithological discrimination *see* rock discrimination
- Local Area Coverage (LAC), 58
- loess, 116, 189
- logarithmic stretch, 128
- logical operators, 222
- look-up tables (LUTs), 125, 128, 129, 130, *Plate 8.1*
- Los Angeles (US), 238
- Louisiana (US), 114
- low-pass filters, 134, 135
- Luliang Shan (China), 189
- luminance, 24
- Lut Block (Iran), 169
- LUTs (look-up tables), 125, 128, 129, 130, *Plate 8.1*
- Mach bands, 24, 25
- magmas, 84, 86, 88, 101, 102
- basaltic, 109
- and volcanism, 107–12, *Plate 9.9*
- magnetite, 10
- magnetic data
- anomalies, 215, 217, *Plates 8.1–2, Plate 9.6*
- in hydrocarbon exploration, 239
- pseudocolours, 216
- magnetic tape, data storage, 39
- Makran Range (Iran), 101, 169
- Mali, 185
- manned spacecraft
- cameras, 55
- data, 55–7
- disadvantages, 54
- missions, 55
- Space Shuttle, 55–7
- maps
- from images, 206–7
- geomorphological, 229–30
- hydrological, 223–4
- production standards, 221
- projections, 253–4
- reconnaissance, 233
- rectification, 254
- structural, 228
- surface materials, 151–2
- symbolology, 208
- terrestrial, 232
- see also* contour maps; geological mapping; geomorphological mapping; mineral mapping; spectral mapping
- Marathon Basin (US), 107
- marbles, *Plate 9.1*
- marine erosion, 69, 75
- Mars, 177, 232
- geomorphology, 229
- remote sensing, 6
- masking, 221
- Mauna Loa (Hawaii), 109
- maximum-likelihood classification, 149–51, *Plate 5.9*
- meanders, 72, 113
- median filters, 143, 257
- mediasats, 66
- medium-pass filters, 134
- memory, and human vision, 19
- Mercator projections, 30
- Mercury programme, 55
- Mersa Fatma (Eritrea), *Plate 1.1*
- mesas, 77
- metal exploration, 236–7
- drainage surveys, 211–12
- geobotanical methods, 237
- geochemical data, 236–7
- and geographical information systems, 237
- geophysical data, 236
- targeting, 237, *Plate 9.6*
- metal-hydroxyl bonds, 9
- metals, dielectric constant, 176
- metamorphic rocks, 91, *Plate 5.1, Plate 6.2, Plate 8.3, Plate 8.4*
- cleavage, 90
- erosion resistance, 90
- folding, 90
- radar images, 186, 193
- recognition, 88–90
- metamorphism, mechanisms, 88
- metapelites, *Plate 8.4*
- metasediments, *Plate 9.2–3*
- meteorite impacts, 103
- meteorological satellites (metsats), 57–9, 66
- resolution, 57
- Meteosat, 57
- methane, 6, 160
- Metric Camera, 56
- Mexico City, 242
- Mg-OH bonds, 9, 142, 152
- microwave imaging, 43–8
- resolution, 43
- see also* passive microwave imaging; radar images
- bands, 43
- detection, 43
- emittance, 43, 176
- generation, 2, 43
- microwave systems
- active, 43
- instantaneous field of view, 43
- passive, 43
- see also* radar
- Mid-Atlantic Ridge, 231
- Middle East, 115
- mid-infrared (MIR), 4, 7
- Mie scattering, 7
- mineralization, 219, 220, 222–3, 244, *Plate 9.6*
- models, 237
- mineral mapping
- and classification, 151–2
- and ratioing, 145
- spectral, 154
- subsurface, 206
- minerals
- band choice, 130
- detection, *Plate 3.2, Plate 9.6*
- ferromagnesian, 84, 236
- spectra, 7–11
- see also* carbonate minerals; clay minerals; hydroxyl-bearing minerals; iron minerals
- minimum-distance-to-means classification, 149, 150
- minus-blue panchromatic images, 34
- Miranda (moon), 232
- MIR (mid-infrared), 4, 7
- missiles, guidance systems, 214
- Mississippi (US), 114
- Moderate Resolution Imaging Spectrometer (MODIS), 66
- Modular Optical Electronic Multispectral Scanner (MOMS), 56, 249
- modulation, 22
- modulation transfer function (MTF), 22–3, 134
- achromatic, 23, 253
- chromatic, 23
- moisture content
- and radar images, 177, 188
- and thermal images, 166
- Mojave Desert (US), 167, 172, 193, *Plate 6.3*
- MOMS (Modular Optical Electronic Multispectral Scanner), 56, 249
- Mongolia, 190
- monoclines, 92
- montmorillonite, 9
- Moon
- craters, 102–3, 232
- gamma-ray surveys, 50
- geomorphology, 229
- optical illusions, 30
- moraines, 116, 117, 119
- Morocco, 106, *Plate 6.2*
- mountains, formation, 229
- Mount Etna (Italy), 243, *Plate 9.9*
- Mount Fuji (Japan), 65
- Mount St Helens (US), 111, 112, 243
- Mount Tom Price (Australia), 182
- Mount Turner (Australia), 182
- moving averages, in gridding, 210
- MS (multiple sclerosis), 243
- MSS *see* Multispectral Scanner (MSS)
- MTF *see* modulation transfer function (MTF)
- mud avalanches, 112
- mudflows, 110, 112, 241, 243
- mudstones, 77, 78, 80, 90
- thermal inertia, 165
- multiple sclerosis (MS), 243
- multispectral data, histograms, 123–5
- Multispectral Scanner (MSS), 61, 62, 68, 139, 141
- false-colour limitations, 129, *Plate 3.3*
- thermal images, 168–9
- multispectral thermal data, 172–4
- multitemporal analysis, 156–7
- multivariate data, in geographical information systems, 206–7
- multivariate raster images, 218–21
- Munsell, Albert Henry (1858–1918), colour atlas, 28
- muscovite, 9, 10
- mylonite, 99
- NASA *see* National Aeronautics and Space Administration (NASA)
- NASDA (National Space Development Agency (Japan)), 62
- National Aeronautics and Space Administration (NASA), 52, 53, 56, 57, 204
- Landsat series, 59–62
- missions, 55, 58, 66
- National Intelligence Mapping Agency (NIMA) (US), 213, 214, 229
- National Oceanic and Atmospheric Administration (NOAA)
- orbital plans, 60
- satellites, 57, 58, 157, 166–7
- National Space Development Agency (NASDA) (Japan), 62
- natural gas *see* hydrocarbon resources
- neighbourhood functions, in geographical information systems, 223
- nerve cells, 17, 18–19
- neural networks, 153

- Nevada (US), *Plate 6.4, Plate 9.9*
 Newton's fringes, 203, 204
 New York (US), 114, 118
 nickel, anomalies, 219
 Nigeria, 89
 night vision, 18
 Nile, River, 179
 NIMA (National Intelligence Mapping Agency (US)), 213, 214, 229
 Nimbus series (satellites), 58, 59, 60
 NOAA *see* National Oceanic and Atmospheric Administration (NOAA)
 noise, random, 143, 257
 non-image data, 206–26
 forms of, 207–8
 hue–intensity–saturation transforms, 216, 219–20
 raster format, 208–21
 in stereomodels, 220–1
 vector format, 208
 nonlinear stretch, 128
 nonselective scattering, 7
 nonsilicates, spectra, 10, 11
 North Africa, *Plate 3.4*
 North America, *Plate 8.2*
 North Dakota (US), 117
 NOT operator, 222
 Nottinghamshire (UK), 41
 Nova Scotia (Canada), *Plate 9.3*
 nuées ardentes, 110–12

 obsequent streams, 70, 93
 oceanographic satellites, 57–9, 66
 oceans
 floor topography, 214, 215
 landforms in, 231
 surface elevation, 214
 offset streams, 100, 101
 O–H bonds, 9
 oil exploration *see* hydrocarbon exploration
 Oklahoma (US), 94, 166
 olivine, 8, 10
 Oman, 87, 90, 98, 101, *Plate 5.9*
 onion-skin weathering, 88
 ophiolites, 90, 101, *Plate 5.7, Plate 5.9*
 OPS (Optical Sensor), 62
 optical illusions, 19
 in aerial photographs, 36–7
 contours, 206
 and image interpretation, 30–1, 96
 Mach-band, 24
 Optical Sensor (OPS), 62
 optic nerve, 17
 orbiting satellites
 advantages, 51
 characteristics, 53–4
 disadvantages, 51
 spectral bands, 58
 see also geostationary satellites; Sun-synchronous satellites
 orbits
 geosynchronous, 53–4
 polar, 54
 ores, detection, 152, 236–7
 OR operator, 222
 orthoclase, 10
 orthophotographs, 255
 outcrops, 78, 86, 148, 154, *Plate 4.2*
 displacements, 99
 outwash deposits, 119
 overlay functions, in geographical information systems, 222–3

 overstep, 94
 overtones, 4
 oxbow lakes, 113, 114
 oxygen, 6
 energy level, 7
 ozone, 6

 Pakistan, 59, 101, 187
 Pamir Range (Asia), 187
 panchromatic film, 34, 56
 parallax, 29, 30, 37
 and digital elevation models, 220, 233
 distortions, 254, 255
 and elevation data, 214
 and radar images, 180, 197
 and stereometry, 249–52
 parallax bars, 251
 parallax wedges, 251
 parallelepiped classification, 149, 150
 particle-wave duality, 19
 passive microwave imaging, 176
 resolution, 43
 passive remote sensing, 4
 pathfinder elements, in geochemical exploration, 219
 pattern, 68–9
 pattern recognition, 145–57
 and artificial intelligence, 145, 147
 spatial, 155–6
 see also classification
 PCA *see* principal component analysis (PCA)
 PDF (probability density function) *see* histograms
 peat, health risks, 243
 peneplains, 72
 Pennsylvania (US), 181
 perceptual ability, and age, 16
 peridotites, 90, *Plate 5.1*
 raster images, 219
 thermal inertia, 165
 permafrost, 120
 perspective views, *Plate 9.2*
 petroleum exploration *see* hydrocarbon exploration
 pH, and metal exploration, 236
 photoelectronic detectors, 39
 photogeology, 68–121, 180
 photographic film *see* film
 photographs (aerial) *see* aerial photographs
 photography, 32–7
 colour, 34–6, 68
 daguerreotype, 50
 infrared, 34, 50
 minus-blue, 34
 oblique, 36
 vertical, 36, 37
 vignetting, 37
 see also aerial photography; colour photography
 photomultipliers, 50
 photons, 1
 photosensitive detectors, 38
 photosynthesis, catalysis, 12
 phreatophytes, 241
 physical resources, exploration, 234
 piecewise stretch, 128
 pigeonite, 8
 pinnate drainage, 116
 Pissgah Crater (US), 167, 172, *Plate 6.3*
 Pittsburgh (US), 181, 199

 pixels, 32, 40–1, 61, 122
 classification, 151
 distortions, 254
 histograms, 123
 purity, 155
 ratioing, 143
 spectra, 155
 Planck's law, 2
 planetology, 229, 232
 plankton blooms, 239
 plant cells, structure, 12
 plants
 anomalies, 237
 constituents, 12–13
 dielectric constant, 177
 energy absorption, 12
 hydrocarbon contamination, 239
 nutrients, 80–1
 roots, 241
 spectra, 144
 see also trees; vegetation
 plateau basalts, 85
 point-spread function (PSF), 21–2, 33
 polarized radiation, 1–2, 180
 pollution, 66
 polynomial transforms, 254
 Pompeii (Italy), 243
 pores, water-containing, 14
 porphyry, 237
 mineralization, 222–3
 potassium isotopes, 219, 220
 power, 2
 exitance, 2
 irradiance, 2
 radiance, 2
 spectral, 2
 radiant flux, 2
 spectral, 2
 radiant flux density, 2
 emittance, 2
 precipitation, and water supplies, 240
 primary colours
 additive, 26, 34–5, 219, *Plate 2.1*
 subtractive, 26, 35
 primitives, 155
 principal component analysis (PCA), 138–43
 canonical, 141, 152
 in change detection, 157
 and classification, 151
 covariance matrices, 139
 decorrelation, 138–9, 140
 directed, 141–2
 eigenvalues, 139–41, 153
 eigenvectors, 139–41, 142, 153
 in image processing, *Plates 5.4–5, Plates 6.3–4, Plate 7.4*
 in spectral mapping, 153
 in thermal image processing, 169
 principal point, 36, 37
 probability density function (PDF), 123
 projections, 253–4
 proximity analysis, 223
 pseudocolours, in raster-format data, 215–16
 pseudoshadowing, 137, 216
 PSF (point-spread function), 21–2, 33
 psychophysiology, and image interpretation, 27
 pupil
 diameter, 17
 dilated, 21

- pushbroom systems, 42–3, 56, 249, 253
 advantages, 43
 aerial, 52
 arrangement, 42
 resolution
 spatial, 42
 spectral, 43
 signal-to-noise ratio, 43
 pyrite, 145, 236
 pyroclastic rocks, 84, 85
 thermal inertia, 165
- Qaidam Basin (China), *Plate 4.4*
 Qinhai Province (China), *Plate 4.4*
 quanta, 1
 quantum energy, 2
 quartz, 9, 10, 84, 85–6
 cementation, 77
 spectra, 10, 237
 thermal inertia, 165, 171
 Quebec (Canada), 103
- RADAM project (Brazil), 53, 230
 radar, 43–8
 airborne, 53
 altimetry, 214, 231, 238
 antennae, 43, 46
 beam width, 44–6
 blooms, 180–6
 coherence, 176
 corner reflectors, 177, 180–6, 227
 depression angles, 43, 47, 48, 176, 178, 195–7
 development, 43, 50
 elevation data, 214
 flatirons, 47
 ground range, 43
 incidence angles, 47
 letter codes, 43
 look direction, 189–95
 parallax, 197
 penetration, 176, 179, 189
 polarization, 47, 180
 pulse length, 44
 range, 43
 real-aperture, 46
 in remote sensing, 4
 resolution, 44–5
 azimuth, 45
 ultrafine, 51
 sideways-looking, 43, 44, 50, 231
 slant range, 43, 44, 45, 47, 48
 specular reflection, 177
 surface material interactions, 176–80
 texture analysis, 199–200, 201–2, 203, 227
 see also backscatter (radar); synthetic-aperture radar (SAR)
 radar images, 43–8, 176–205, 230
 advantages, 176, 195
 combined data, 202–3
 and depression angles, 176, 178, 180, 195–7
 digital processing, 199–202
 foreshortening, 48, 49
 geological features, 180–9
 interpretation, 180–203
 layover, 47, 48
 and look direction, 189–95
 multifrequency, 180, 198–9
 multipolarized, 180, 197–8
 and parallax, 180, 197
 range compression, 47
 shadows, 48, *Plate 7.1*
 side-lighting, 48
 stereoscopic, 48, 180, 197
 topographic accentuation, 180, 197
 topographical effects, 180–9
 radar interferometry, 203–4, 213–14, 242
 Radarsat, 64, 197, 230–1
 radar scattering coefficient, 176, 178
 radar scatterometry, 241
 radial drainage, 72, 73
 radial relief displacement, 36, 37
 radiance, 2, 154, 161
 radiant flux density, 2
 radiant temperature, 11
 radiation-matter interactions, 161
 radioactivity, health risks, 244
 radon, health risks, 243
 ramp stretch, 128
 raster-format data, 122, 207–8, 225
 non-image, 208–21
 processing, 215–18
 pseudocolours, 215–16
 see also multivariate raster images
 rasters, 19, 32, 40, 122
 raster-vector conversions, 221
 ratioing, 143–5
 disadvantages, 144
 and geobotany, 145
 and geochemical data, 219
 and mineral mapping, 145
 Rayleigh criterion, 177–8, 197, 198, 199
 Rayleigh scattering, 6–7, 13
 RBV (Return-beam Vidicon), 61, 68
 receptors
 cones, 17–18, 21, 26–7
 density, 18
 rods, 17–18, 21, 24
 rectangular drainage, 72, 73
 rectification, 254
 red edge, 12
 red-green-blue (RGB) images, 199, 233, 237, *Plates 5.3–5*
 band choice, 129, *Plate 5.1*
 contrast stretching, 131–3
 decorrelation stretching, 142–3
 multivariate rasters, 218–21
 Red Sea, *Plate 1.1*
 Red Sea Hills (Sudan), 88, 105, 193, 241
 redundancy
 spatial, 19–20
 spectral, 131
 tonal, 19–20
 reflectance, 5
 spectral, 138, 147, 148, 154, 161
 reflection
 diffuse, 5–6
 Lambertian, 5
 seismic, 238
 specular, 5, 177
 reflection spectra, 5, 7–13
 registration, 122, 233
 in geographical information systems, 221
 regression analysis, in gridding, 209
 rejuvenation, 70, 72
 remote sensing
 atmospheric effects, 6–7
 criticisms, 96
 economics, 235
 and field observation, 227
 future prospects, 65–7
 historical background, 50–1
 spectral range, 27
 strategies, 227–8
 note sensing missions
 future prospects, 65–7
 manned, 55–7
 unmanned, 57–64
 remote sensing systems
 active, 4
 constraints, 7
 energy detection, 4–5
 passive, 4
 resampling, 254
 cubic convolution, 254
 nearest-neighbour, 254
 resequent streams, 70, 93
 resolution
 film, 33
 radiometric, 39
 spatial, 19–24, 39, 42, 227–8
 spectral, 39, 42
 Resurs (satellites), 64
 retina, 16, 29
 cross-section, 17
 fovea, 17, 18, 24, 29
 nerve cells, 17, 18–19
 receptors, 17–18
 retinal frequency, 23
 Return-beam Vidicon (RBV), 61, 68
 RGB images see red-green-blue (RGB) images
 rhodopsin, 17
 rills, 72
 ringing, 138
 risk assessment
 in exploration, 234
 geochemical data, 243–4
 subsidence, 241–2, 243
 volcanoes, 243
 roches moutonnées, 74
 rock discrimination, 227, *Plate 9.1*
 colour rendition, 219
 general approach, 76
 Rocklea Dome (Australia), 182
 rocks
 colour, 77–8, 81, 83, 84
 dielectric constant, 176–7
 hydrothermal alteration, *Plate 9.5*
 sampling, 211
 thermal inertia, 165, 171
 thermal properties, 162
 ultramafic, 219
 volcanic, *Plate 5.9*
 wet, 14
 see also igneous rocks; metamorphic rocks; pyroclastic rocks; sedimentary rocks
 rock spectra, reflection, 7–11
 rock types
 and erosion, 69
 recognition, 76–90
 and water exploration, 240–1
 rods, 17–18, 21, 24
 Roraima Formation (Venezuela), 198
 rotational transitions, 4
 roughness
 and backscatter (radar), 177–9, 187, 189
 surface materials, 177–9, 217, *Plate 7.3*
 rubidium, concentration, *Plate 8.3*
 Russia, 59
 sag ponds, 100
 Sahara Desert, 179, 192, 196
 salt diapirs, 102, 103
 Salt Range (Asia), 187
 salts, 81
 salts flats, 124

- sand dunes, 114–16, 190
- sandstones, 77, 79, 84, *Plates 6.1–2*
 - and granites compared, 86–8
 - isotopes, 219
 - radar images, 200
 - thermal inertia, 165, 171
- San Rafael Swell (US), 170, *Plate 6.1, Plate 7.4*
- Santa Barbara (US), 194
- Santa Ynez Mountains (US), 194
- saponite, 9
- SAR *see* synthetic-aperture radar (SAR)
- satellite images, resolution, 51
- satellites
 - geostationary, 53–4
 - military, 50–1
 - oceanographic, 57–9, 66
 - orbits, 53–4
 - tracking and data relay, 61
 - see also* meteorological satellites (metsats); orbiting satellites; Sun-synchronous satellites
- saturation, 125, 126, 132
- Saturn-V launch vehicle, 55
- scabland, 119
- scale, 69, 227–8
- scarps, 100
- scattering
 - Mie, 7
 - nonselective, 7
 - Rayleigh, 6–7, 13
 - in water, 13
 - see also* backscatter (radar)
- scheelite, *Plate 9.5*
- Scotland, 74, 99, 168, *Plate 8.3*
- sea level, factors affecting, 214
- Seasat, 55
 - radar altimeters, 214, 231
 - radar images, 195, 197
 - SAR system, 58–9
- seasonal images, 69, 156, *Plate 5.10*
- Sea Wide-field Sensor (Sea-WIFS), 58
- second derivative filters, 136
- Second World War, aerial photography, 50, 51
- sedimentary basins, structure, 238
- sedimentary rocks
 - erosion resistance, 76–7
 - and igneous rocks, 83–4
 - permeability, 76–7
 - porosity, 69, 76
 - recognition, 76–83
 - thermal inertia, 171, 173
 - thermal spectra, 173, 174
 - vegetation cover, 77
- sediments
 - chemically precipitated, 81–3
 - groundwater, 241
 - lacustrine, *Plate 4.4*
 - marine, 187
 - morphology, 76–7
 - outcrops, *Plate 4.2*
 - radar penetration, 179
 - raster images, 219
 - sampling, 211
 - suspended, 13
 - thermal inertia, 165
 - see also* clastic sediments
- seepages, hydrocarbon resources, 238–9, *Plate 9.7*
- seif dunes, 115
- seismic events
 - monitoring, 99–100
 - risk assessment, 242
- seismic reflection, and hydrocarbon exploration, 238
- Selby Coalfield (UK), *Plate 9.9*
- selective radiators, 10
- self-organising mappers, 153
- shadows
 - optical illusions, 30
 - radar images, 48, *Plate 7.1*
- shales, 78, *Plate 9.5*
 - health risks, 243
 - radar images, 200
 - reflectance, 147
- Shanxi Province (China), 189
- shapes, 69
 - optical illusions, 31
- shear zones, 104, 105
- shield volcanoes, 109–10
- short-wavelength infrared (SWIR), 7, 9, 236
- shutter ridges, 100
- Shuttle Imaging Radar (SIR)
 - SIR-A, 55, 195, 197
 - SIR-B, 56, 66
 - SIR-C/XSAR, 56, 66, 204
- Shuttle Multispectral Infrared Radiometer (SMIRR), 55
- Shuttle Radar Topography Mission (SRTM), 66, 204, 214, 229
- siderite, 10
- sideways-looking radar, 43, 44, 50, 231
- sight *see* human vision
- signal-to-noise (S/N) ratio, line scanners, 39
- silicates, 236
 - spectra, 10, 11, 172
 - structure, 9
 - thermal spectra, 174
- silicon, energy level, 7
- silicon dioxide, in igneous rocks, 174
- sills, 85, 95
- siltstones, 77, 78, 80, 90, 192, *Plate 6.4*
 - radar images, 200
 - thermal inertia, 165
- sinkholes, 83
- Si–O bonds, 10, 160, 172
- SIR *see* Shuttle Imaging Radar (SIR)
- Skylab, 55
- skylight, 144
- slopes
 - stability assessment, 241
 - topographic, 213
 - types of, 91–2
- SMIRR (Shuttle Multispectral Infrared Radiometer), 55
- snow
 - radar images, 186, 187
 - spectra, 14
- S/N ratio (signal-to-noise ratio), 39
- Socompa (Chile), 111
- soils
 - anomalies, 14
 - dielectric constant, 176–7
 - erosion, 240
 - formation, 77, 80–1
 - geochemical surveys, 243–4
 - roughness, 179
 - sampling, 211
 - thermal inertia, 165
- solar energy, 154
 - absorption, 6
- solar radiation
 - absorption, 161
 - diurnal variations, 164
 - and remote sensing, 4
 - surface heating, 161
- solar shadowing, 164, 165, 216, 223
- Somalia, 81
- sonar images, 231
- sour gas, formation, 239
- South Africa, *Plate 5.10*
- Space Imaging, 66
- Space Shuttle, 54
 - antennas, 204
 - Challenger disaster, 57, 66
 - data, 55–7
 - imaging systems, 56, 229
 - remote-sensing experiments, 57
 - topographic surveys, 213–14
- spatial acuity, 23–4
- spatial filters, 134–8
- spatial-frequency filtering, 133–8, 200–1
 - definition, 134
 - scale issues, 134
- spatter cones, 110
- specific heat, 162
- spectra
 - absorption, 5
 - emission, 5
 - reflection, 5, 7–13
 - transmission, 5
- spectral coverage, and geological applications, 228
- spectral mapping, 52–5
 - neural-network techniques, 153
 - self-organising, 153
- spectral radiance, 2
- spectral radiant flux, 2
- spectrometers, 66, 141
 - gamma-ray, 49–50, 219
 - imaging, 49
- spectroradiometers, 49, 57, 152
 - flat-field calibration, 154
 - output spectra, *Plate 3.2*
- specular reflection, 5, 177
- spessartine (garnet), 8
- splines, 217
 - cubic, 209, 210
 - in gridding, 209–10
- SPOT *see* Système Probatoire de l'Observation de la Terre (SPOT)
- sprinklers, radar images, 195
- spy planes, 50
- square filters, 138
- SRTM (Shuttle Radar Topography Mission), 66, 204, 214, 229
- staffed spacecraft *see* manned spacecraft
- staurolite, 8
- Stefan-Boltzmann constant, 3
- Stefan-Boltzmann Law, 3, 9, 11, 58, 154, 160, 176, 243
- stereometers, 251
- stereometry, 249–52
- stereomodels, 29, 220–1
- stereopsis, methods, 249–50
- stereoptic viewing, 29–30, 249–50
 - anaglyphs, 250
 - vertical exaggeration, 250–1
- stereoscopes, 68, 250, 252
- stereoscopic images
 - aerial photographs, 37, 68–121
 - and dip estimation, 93
 - elevation data from, 214
 - engineering applications, 241
 - from non-image data, 220–1
 - in geomorphology, 229
 - radar, 48, 180, 197
- stereoscopic viewing, 68–121
 - in geological mapping, 233

- stereoscopic vision, 19
 stomach cancer, 243
 strata, 76
 horizontal, 77, 78
 tilted, 77
 stratigraphic relationships, 90–5
 stratovolcanoes, 108, 109, 188
 eruption kinetics, 110–12
 streams
 braided, 113, 119
 catchments, 211, 212, 223
 classification, 71
 consequent, 70, 93
 drainage patterns, 70–3
 insequent, 70
 obsequent, 70, 93
 offset, 100, 101
 order, 211–12, 223
 resequent, 70, 93
 sampling from, 211–12
 subsequent, 70, 72
 types of, 70
 strike, 91–3, 213, 233
 strike cut-offs, 97
 striping, 253
 image, 255
 strontium, concentration, *Plate 8.3*
 structural mapping, 228
 structural relationships, 95–107
 subsequent streams, 70, 72
 subsidence
 detection, 243
 mechanisms, 242–3
 risk assessment, 241–2
 subsurface dams, 240–1
 subsurface data, 206
 subtractive primary colours, 26, 35
 Sudan, *Plate 4.5*
 Red Sea Hills, 88, 105, 193, 241
 Sudan–Eritrea border, 126, 131, 135, 136,
 138, 140, 146, *Plate 5.1*, *Plates 5.3–6*,
 Plate 5.8
 Suleiman Range (Pakistan), 101
 sulphides, 81, 236–7, 244
 Sun
 illumination effects, 143–4, 151, 189–95
 radiation, 4
 surface temperature, 3
 sunsets, and Mie scattering, 7
 Sun-synchronous satellites, 58, 64, 171
 advantages, 57
 disadvantages, 54
 groundtracks, 54
 resolution, 54
 see also Landsat
 superficial deposits, 107–20
 superimposed drainage, 70
 superposition, 93–4
 supervised classification, 151, 153–4
 surface dams, 240
 surface drainage, patterns, 71
 surface geology, data, 206
 surface heat balance, 161–2
 surface materials
 backscatter (radar), 176
 classification, 147–52, 169
 dielectric constant, 176–7, *Plate 7.3*
 maps, 151–2, 154
 radar interactions, 176–80
 radiant emission, 160
 roughness, 176, 177–9, 217, *Plate 7.3*
 thermal inertia, 165, 171
 thermal properties, 160, 162–3, 171
 surfaces, reflectance, 143–4
 surface temperature, 161–2
 diurnal variations, 163–4, 169
 Earth, 4, 11, 43, 160
 profiles, 163–4
 Sun, 3
 surface water, 240
 surveys
 aerial, 51
 drainage, 211–12, 244
 geochemical, 243–4
 geophysical data, 210
 line, 210
 SWIR (short-wavelength infrared), 7, 9, 236
 synclines, 104
 synthetic-aperture radar (SAR), 46
 applications, 53, 58–9, 64, 230–1
 azimuth resolution, 46
 depression angles, 66
 Doppler shift, 46
 holograms, 46, 47
 images, 65, 176
 Syria, 191
 Système Probatoire de l'Observation de la
 Terre (SPOT), 157
 data, 62
 false-colour limitations, 129
 ground stations, 62
 sensors, 62–3
 spectral bands, 58
 stereoscopic images, 214, 229
 stereoscopic potential, 62, 249
 Tabriz (Iran), 188
 tar *see* hydrocarbon resources
 targeting, 234–6, 237, *Plate 9.6*
 TDRs (tracking and data relay satellites),
 61
 tectonics, 95–107
 telemetry, line scanners, 39
 television
 additive primary colours, 26, 34–5
 colour transmission, 27
 Television and Infrared Observation
 Satellite (TIROS-1), 57, 60
 Television and Infrared Observation
 Satellite (TIROS-N), 157, 166–7
 terraces, 114
 terrain analysis, 228, 230
 Terra (satellite), 66
 Texas (US), 107
 texture, 68
 image variance, 155
 radar data analysis, 199–200, 201–2, 203,
 227
 Thematic Mapper (TM)
 applications, 237
 characteristics, 61–2
 false-colour images, *Plate 3.3*
 limitations, 129–30
 image size, 131
 ratio images, 145
 reflected bands, 138
 thermal images, 166
 wavebands, 61, 142, 153
 thermal capacity, 162, 163
 thermal conductivity, 162
 thermal cross-over, and thermal images,
 164–5
 thermal data
 in classification, 171
 see also multispectral thermal data
 thermal diffusivity, 162
 thermal images, 160–75, 230
 analogue, 160
 day-time, 164, 168, 170, 171, 172, *Plates*
 6.1–3
 detectors, 160
 digital, 160
 false-colour, 168–9
 grey-tone, 164–7
 interpretation, 160, 162
 qualitative, 164–9
 semiquantitative, 160, 169–71
 night-time, 164, 165, 168, 170, 171, 172,
 Plates 6.1–3
 small-scale low resolution, 166–7
 and thermal cross-over, 164–5
 topographic effects, 164, 165
 see also infrared images
 thermal inertia, 11, 162, 166, 167, 173, *Plate*
 6.1
 estimation, 160, 163, 167, 169–71
 surface materials, 165
 and surface temperature, 163–4
 see also apparent thermal inertia (ATI)
 Thermal Infrared Multispectral Scanner
 (TIMs), 57, 130, 173
 thorium isotopes, 219, 220
 thrust tectonics, 94
 Tibet, 59
 Tibetan Plateau, 96, 102, *Plate 4.3*
 tides, 214
 till, 116, 117–18
 tilting
 faults, 100
 unconformities, 95
 TIMS (Thermal Infrared Multispectral
 Scanner), 57, 130, 173
 tin, anomalies, 237, *Plate 9.6*
 TIN (triangular irregular network), 213, 252
 TIROS-1 (Television and Infrared
 Observation Satellite-1), 57, 60
 TIROS-N (Television and Infrared
 Observation Satellite-N), 157, 166–7
 TM *see* Thematic Mapper (TM)
 tone, 68
 topography
 effects
 on radar images, 180–9
 on thermal images, 164, 165
 elevation data, 213–14
 visualization, 216
 Toulouse (France), 62
 Tournachon, Gaspard-Félix (1820–1910), 50
 tracking and data relay satellites (TDRs),
 61
 training areas, 149, 153–4
 training sets, 141–2
 transition metals, spectra, 8
 transitions
 bond-bending, 9
 bond-stretching, 9
 charge-transfer, 8
 electronic, 4, 8
 rotational, 4
 vibrational, 4
 transmission spectra, 5
 transmittance, 5, 161
 water, 13
 transpiration, 13
 Transvaal (South Africa), *Plate 5.10*
 transverse dunes, 115
 trees, *Plate 9.7*
 false-colour images, *Plate 3.1*
 spectra, 13

- trellis drainage, 72–3, 77, 80
 trend-surface analysis
 in gridding, 209
 regional fields, 217
 residual fields, 217–18
 triangular irregular network (TIN), 213, 252
 tristimulus colour theory, 28
 truncated spurs, 100
 tuffs, 84, 110, *Plate 6.4*
 Turkey, 242

 U2 (spy plane), 50
 ultramafic rocks, raster images, 219
 ultraviolet radiation, generation, 2
 uncertainty principle, Heisenberg's, 19
 unconformities, 93, 94–5, 233, *Plate 4.2*, *Plate 9.2*
 folding, 95
 tilting, 95
 United Kingdom (UK) *see* England; Scotland
 United States (US), 86, 108, 115
 Appalachian Plateau, 55, 59, 181, 199, *Plate 4.1*
 Arizona, 70, 86, 108, 115, *Plate 4.1*
 California, 101, 194, 238, *Plate 5.7*, *Plate 9.9*
 Mojave Desert, 167, 172, 193, *Plate 6.3*
 Colorado, 165, 168
 Colorado River, 70, *Plate 4.1*
 Goddard Space Flight Center, 55, 57
 Grand Canyon, 70, *Plate 4.1*
 Indiana, 117
 Kansas, 78
 Kentucky, 55, 59, 82, 91
 Louisiana, 114
 Mississippi, 114
 Mount St Helens, 111, 112, 243
 Nevada, *Plate 6.4*, *Plate 9.9*
 New York, 114, 118
 North Dakota, 117
 Oklahoma, 94, 166
 Pennsylvania, 181
 Pittsburgh, 181, 199
 Texas, 107
 Utah, 71, 79, 80, 100
 San Rafael Swell, 170, *Plate 6.1*, *Plate 7.4*
 Virginia, 55, 59, 80
 Washington, 116
 Wisconsin, 119
 Wyoming, 89, 92, 202, *Plate 7.3*
 Wind River Basin, 200, *Plate 6.5*, *Plate 7.1*
 United States Geological Survey, 53
 missions, 55
 Universal Transverse Mercator (UTM)
 projection, 254
 unmanned spacecraft
 advantages, 57
 data, 57–64
 historical background, 32
 unstaffed spacecraft *see* unmanned spacecraft
 unsupervised classification, 151
 uranium
 health risks, 243
 isotopes, 219, 220
 Uranus, 232
 Utah (US), 71, 79, 80, 100
 San Rafael Swell, 170, *Plate 6.1*, *Plate 7.4*
 surface materials, 148

 UTM projection (Universal Transverse Mercator projection), 254

 valley glaciation, 73–4
 valley shape, 72
 U-shaped, 74, 116, 119
 V-shaped, 70, 93
 valley V's, 93, 105, 106, 107
 variance, 139, 155
 vector-format data, 207–8, 221, 225
 non-image, 208
 vegetation
 anomalies, 237, 239
 band choice, 130
 classification, 148–9, 241
 dielectric constant, 177
 and fault detection, 97
 in flood plains, 113
 geological control, 77, 88, 133, *Plate 5.10*
 image processing, 124, 141–2, 228
 monitoring, 61
 multitemporal analysis, 156
 patterns, 120
 in radar images, 184, 198–9
 in ratio images, 144
 recognition, 76, 77
 red edge, 12
 reflectance, 147
 reflectivity, 34
 roughness, 179
 and slope stability assessment, 241
 spectra, 12–13, 144
 and springs, *Plate 9.8*
 stressed, 145
 on thermal images, 165–6
 thermal inertia, 164, 165
 in water exploration, 240, 241
 see also plants
 Venezuela, 198
 Venn diagrams, 222
 Ventura Basin (US), 238
 Venus, 177, 232
 very-near infrared (VNIR), 7, 12, 34, 236–7
 Vesuvius (Italy), 243
 vibrational transitions
 bond-bending, 4
 bond-stretching, 4
 video cameras, 32
 Vidicon cameras, 37–8
 colour, 38
 see also Return-beam Vidicon (RBV)
 vignetting, 37
 Virginia (US), 55, 59, 80
 vision (human) *see* human vision
 visual cortex, 16–19, 29, 31
 cells, 224
 complex, 19
 hypercomplex, 19
 simple, 19
 visual perception, tonal limitations, 26
 VNIR (very-near infrared), 7, 12, 34, 236–7
 Vogelkop Region (Iran), 184, 186
 volcanic rocks, *Plate 5.9*
 volcanoes, 107–12, 231, 232, *Plate 7.2*, *Plate 9.9*
 cones, 84, 102, 103, 107–9, 186
 eruption detection, 243
 eruption kinetics, 110–12
 explosive activity, 110

 gases, 110
 parasitic, 188
 radar images, 188, 191, 199, 201
 risk assessment, 243
 shield, 109–10
 thermal images, 167
 see also lavas; stratovolcanoes

 Wamsutter Arch (US), 202, *Plate 7.3*
 Washington (US), 116
 waste disposal
 environmental hazards, 243
 sites, 242
 water, 160
 as blackbody, 13–14
 colour, 13
 depth estimation, 13
 dielectric constant, 176, 177
 economic importance, 234
 light penetration, 13
 light scattering, 13
 molecular, 14
 pore, 14
 reflectance, 147
 spectra, 13–14
 surface, 240
 thermal inertia, 164
 thermal properties, 162
 transmittance, 13
 see also groundwater
 water exploration, 239–41
 and rock types, 240–1
 watersheds, 223, 224
 water supplies
 pollution, 240
 strategies, 240
 water vapour, 6
 wavelength, 1
 wave propagation, 1
 weathering, 69, 85
 carbonates, 82
 onion-skin, 88
 websites, 264–5
 weighting, 151
 West Africa, 34
 Western Europe, *Plate 3.4*
 whiskbroom system, 52
 Wien's Displacement Law, 3, 4, 9, 154, 160, 176
 wind erosion, 69, 75–6
 Wind River Basin (US), 200, *Plate 6.5*, *Plate 7.1*
 Wisconsin (US), 119
 World Meteorological Organization, 57
 World War I, aerial photography, 50, 51
 World War II, aerial photography, 50, 51
 Wyoming (US), 89, 92, 202, *Plate 7.3*
 Wind River Basin, 200, *Plate 6.5*, *Plate 7.1*

 Xinjiang (China), 183
 XOR operator, 222
 x-rays, generation, 2

 Yellow River (China), 189
 Young's modulus, 26
 Young, Thomas (1773–1829), colour theories, 26–7, 218

 Zimbabwe, 87
 zinc, anomalies, *Plate 9.6*